Essays on the Measurement of Economic Expectations

Inaugural-Dissertation
zur Erlangung des Grades Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2008

vorgelegt von
Anna Stangl

Referent: Prof. Dr. Joachim Winter
Korreferent: Prof. Dr. Kai Carstensen
Promotionsabschlussberatung 4. Februar 2009

Acknowledgments

First and foremost I would like to thank my thesis supervisor *Joachim Winter*, who from the very beginning believed in my ideas and gave me the freedom to pursue the chosen projects. I am also very grateful for his teachings on econometrics, behavioral and empirical economics. It repeatedly gave me impetus for my own work.

I am very grateful to *Kai Carstensen* who agreed to serve as second supervisor on my thesis committee. I benefited a lot from his comments to my presentations and papers. I also like to thank *Klaus Abberger*, who has enabled the realization of my research and has always taken the time to discuss it. I am indebted to the *Ifo Institute for Economic Research*, particularly to *Meinhard Knoche* and *Sabine Dehof,* who have allowed the technical implementation of the VAS experiments in the Ifo Business Survey and the Ifo World Economic Survey. I also would like to thank *Paul Kremmel* for helping me with the English language, as well as for encouraging me to apply for conferences and publications. Special thanks go to other colleagues at the *Ifo Institute for Economic Research*, especially to *Dirk Ulbricht, Gabriela Schütz* and to my next room colleague *Klaus Wohlrabe,* with whom I shared the experience of writing a thesis and who readily helped out with books and literature recommendations. My cordial thank goes to *Romy Bonitz*, who as a close friend accompanied me during the years with a lot of friendship and humor.

I thank also participants of the research workshop "Empirical Economics", which was organized every semester by *Joachim Winter*, especially I would like to thank *Florian Heiss* for his vivid discussion and very useful comments. I also owe gratitude to the numerous participants of the conferences at which I had the chance to talk.

A special thank goes to *Oscar Claveria Gonzalez*. As a close friend, he has not only been the one who encouraged me to pursue this research, but also believed in and supported my ideas on the new measurement approach of economic expectations from the very beginning. I will also treasure the experience of the joint work with *Roy Batchelor*, his patience, joyful way of work and the generous readiness to share his insights.

Most importantly I am indebted to my family, my parents and my brother, *Mikhail Stangl*, who encouraged me a lot. I would also like to thank *Florian*, and his parents *Eva* and *Gerhard Wolf*, for love and care they gave me during the last years.

Table of Contents

Chapter 2: Validity of Economic Indicators Derived with the Visual Analog Scale

Chapter 3: Testing the Assumptions of Three-category Based Business Expectations Using the Visual Analog Scale

List of Figures

Chapter 1: Reliability of the Visual Analog Scale for the Measurement of Economic Expectations

Chapter 2: Validity of Economic Indicators Derived with the Visual Analog Scale

Chapter 3: Testing the Assumptions of Three-category Based Business Expectations Using the Visual Analog Scale

List of Tables

Chapter 1: Reliability of the Visual Analog Scale for the Measurement of Economic Expectations

Chapter 2: Validity of Economic Indicators Derived with the Visual Analog Scale

Chapter 3: Testing the Assumptions of Three-category Based Business Expectations Using the Visual Analog Scale

List of Abbreviations

3-Cat.	Three-category scale
ANOVA	Analysis of variance
BCI	Business climate indicator
BTS	Business tendency survey
D	Down
DF	Degree of freedom
EU	European Union
FE	Fixed effects model
GDP	Gross Domestic Product
ICC	Intra-class correlation coefficient
IMF	International Monetary Fond
Kurt.	Kurtosis
Med	Median
MWW	Mann-Whitney-Wilcoxon two-sample statistic / rank sum test
N	Number of observations
NSA	Not seasonally adjusted
Obs.	Observations
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary least squares
PDF	Probability density distribution
PI	Production index
Pr	Probability
Q…	Quarter
S	Same
SA	Seasonally adjusted
Sig.	Significance level
Skew.	Skewness
SPF	Survey of professional forecasters
SSB	Sum of squares between groups
SSW	Sum of squares within groups
Std.	Standard
U	Up
VAS	Visual analog scale
WES	World Economic Survey

Executive Summary

Economic expectations determine economic activity. The accuracy of their measurement is therefore of particular interest to economists. Three general measurement methods of economic expectations have been established over time: (1) qualitative measurement with category-rating scales (e.g., “better”, “unchanged”, “worse”), (2) quantitative measurement by eliciting point forecasts (e.g., 3% output growth) and (3) the measurement of economic expectations eliciting subjective probabilities of particular events (in form of a complete probability distribution of a point forecast). However, these most common measurement methods have serious drawbacks.

- *Current Measurement Methods of Economic Expectations – Strengths and Weaknesses*

The motivation and popularity of using qualitative rather than quantitative variables in surveys has several reasons (OECD, 2003): It is generally much easier for respondents to give qualitative rather than quantitative information and the questionnaire can be completed quicker. Furthermore, qualitative questions are less often a source of inaccuracies or inconsistencies. The idea behind such business confidence questions is that polled experts assess the overall business situation in their company or country by taking into account all the aspects they regard as important. Although no precise information on levels of output, sales, investment, or employment is normally asked, business expectations are used to predict changes in these aggregates in the analysis of business cycles. For these reasons business tendency surveys in the European Union and in the majority of OECD countries are harmonized with regard to the utilization of three-category rating scales in business confidence questions. However, the category-rating scales have serious drawbacks. The data on economic expectations elicited with three-category scales are very limited. The question format allows the respondents to choose only from the three

available options that may not match their real opinions. This leads to a drift towards the central category, which includes responses that have a positive as well as negative tendency but which have not reached a particular threshold, which would lead the respondent to choose one of the extremes. The offering of just three response options results in imprecision and information loss. Furthermore, the distribution delivers almost no information on the dispersion of economic expectations, which is often used as a proxy of uncertainty (Mankiw, Reis and Wolfers, 2003, Doepke and Fritsche, 2006, Giordani and Soderlind, 2003). As a detailed distribution of economic expectations is not known, for the quantification of category responses strong assumption about the distribution and the indifference thresholds that underlie a category assessment are necessary.

Most of the problems with category-rating scales could be solved by eliciting point forecasts, or quantitative information on outstanding orders, profits or turnover. However, in business surveys, companies are, as a rule, reluctant to report quantities, either because of confidentiality issues or time constraints. Qualitative questions, in contrast, usually do not require respondents to consult their accounting records.

Alternatively, firms can be asked to report their subjective probability distribution of future events (Guiso and Parigi, 1999), such as demand, output or profit growth. These kind of questions have several desirable features (Manski, 2004). The standard deviation of an individual histogram is associated with the uncertainty of the individual forecaster. The variance in the aggregate histogram incorporates both individual uncertainty and heterogeneity of expectations. However, this response format also has several drawbacks (Boero et al., 2007, Engelberg et al., 2006, Clements, 2007): Probabilistic questions are time-demanding and tend to cause a high cognitive load to respondents. They are consequently only applicable to people familiar with probability distributions. In surveys of professional forecasters, a tendency was observed that forecasters give more careful consideration to their point predictions than to their probabilistic forecasts (Engelberg et al., 2007). Furthermore, in a variety of surveys researchers observed a general tendency of respondents to be optimistic, i.e., to report high probabilities for positive and low probabilities for

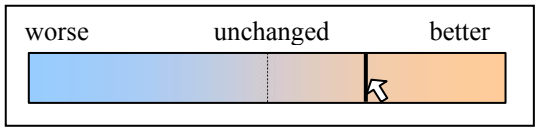
negative events (Guiso, Tiseno and Winter, 2005, Boero et al., 2007, Engelberg et al., 2006, Clements, 2007).

In a nutshell, qualitative questions have many desirable features and are the state of the art in business surveys. Their weakness, however, is the enormous information loss due to the high proportion of neutral responses, lack of information on the distribution of expected changes in the population and strong assumptions about the data generation process. Quantitative questions and subjective probabilities include this information, but have serious practical limitations. This thesis presents a new approach to the measurement of economic expectations, which was made possible by the recent spread of Internet surveys: the visual analog scale (VAS). Although the VAS is a qualitative measurement method, it overcomes most of the problems discussed above.

- *Visual Analog Scales – General Background*

Visual analog scales are rating scales on which a subject ranks the preferences along a continuous line or scale. There are numerous variations of the approach (e.g., length of the line, labels for the ends, vertical or horizontal placement, color etc.).

Visual Analog Scale (VAS)



These scales correspond to the common usage of measurement scales for physical extensions (thermometer, etc.) and are easy to understand and to handle for the respondents. VAS was first described in 1923 by Max Freyd in his article "*The graphic rating scale*" in *The Journal of Educational Psychology* and has become one of the most commonly used measures of feeling and pain intensity in medical research (Jensen et al., 2003). Until recently it has been merely used in personal medical interviews. Application of the VAS in broad-scale surveys has been difficult thus far due to a costly operationalization. With the spread of Internet surveys, the

VAS has become easy to administer. It is a very simple idea, but its benefits compared to other measurement methods of economic expectations are obvious: The VAS enables scores between categories, and the respondent can express not only the direction of his attitude but also its magnitude on a 1-to-100 point scale, which comes close to an interval scale measurement. The scale range allows for a subtle distinction of a respondent's preferences. At the same time, by using a graphic scale with few anchors there is no increase in cognitive load for the respondent, while the information collected is much broader than any conventional category-rating scale would allow. The distributional shape of responses and various measures of dispersion contain additional valuable information on present and future economic activity which can help to detect cyclical turning-points earlier.

- *The Data Sets*

The VAS was implemented in 2005 in two Internet surveys of the Ifo Institute for Economic Research: The monthly business survey in the German manufacturing sector (ifo Konjunkturtest im verarbeitenden Gewerbe) and the quarterly international survey of economists (ifo World Economic Survey). By the date the thesis was completed, 34 monthly data sets containing more than 46,000 VAS responses were collected within the business survey, and 14 quarterly data sets with more than 4,000 VAS responses within the international survey of economists.

The thesis includes three self-contained essays on the new measurement approach of economic expectations. The first two essays are concerned with the estimation of the reliability of the VAS and the validation of economic indicators that can be derived with the VAS. The third essay presents an application of the VAS for the empirical testing of economic theories and assumptions. All essays contain extensive empirical investigations based on micro-data sets both at the firm-level and the level of individual experts.

- *Reliability of the VAS for the Measurement of Economic Expectations*

Chapter 1 aims at the estimation of the reliability of the VAS as a measurement instrument of economic expectations. Reliability refers to the consistency and accuracy of a measurement instrument, and is a necessary condition for validity. According to the definition, reliability is the correspondence of a value measured by a particular scale, with the hypothetically true value. Since the true value is not available, reliability has to be estimated in various ways. There are four general classes of reliability estimates: *Parallel-forms reliability*, *test-retest reliability*, *internal consistency reliability* and *inter-rater reliability*. Each of the classes was addressed. The reliability estimation was conducted on the data of the survey of economists. These data sets enable the estimation of the inter-rater reliability, since a group of economists is surveyed in each country. The main focus of the first essay is not so much on comparing the reliability of the three-category scale with the reliability of the VAS. Given fundamental differences between the two scales with regard to the range of responses, the scales appear not to be perfectly comparable. Although the statistical methods used in the study are theoretically applicable to both instruments, the coarse three-category scale makes the application of most statistical methods used for the estimation of reliability a matter of argument.

The reliability of the VAS is an issue of major interest for several reasons: (1) The thesis presents an unprecedented application of the VAS for the measurement of economic expectations. (2) The VAS has been applied in two self-administered Internet surveys. In the medical surveys, in contrast, the VAS has been merely applied in personal interviews. The Internet environment and the self-administration may negatively affect the scale reliability. (3) The 100-point range of VAS responses clearly overestimates people's discriminatory power relating to the subjects of interest. Consequently, the VAS has an imprecision interval that was necessary to evaluate to be able to estimate a minimum significant difference of VAS scores.

The empirical results of the reliability analysis indicate that the VAS has a very small imprecision interval that is comparable to the imprecision interval estimated in medical studies. Also the estimation results of the other reliability classes were in

favour for the VAS. It was found to be a reliable measurement instrument of economic expectations and in none of reliability classes inferior to the traditional three-category scale.

- *Validity of Economic Indicators Derived with the VAS*

The study in Chapter 2 explores the validity of various indicators derived with the VAS with respect to the economic performance in the German manufacturing sector in the time-period September/2005 to July/2008. As the VAS data allow the estimation of higher moments of distribution, several measures of dispersion, such as the standard deviation, kurtosis and skewness, were evaluated for the presence of a cyclical pattern. Furthermore, VAS data deliver information on the "epistemic" uncertainty, which is contained in the neutral responses of the VAS (Bruine de Bruin, 2000). The dispersion measures of business expectations and the "epistemic" uncertainty measure were found to be closely related to each other, indicating that the dispersion of VAS business expectations contains two components: heterogeneity of business expectations and uncertainty about the future economic development.

Although the number of 34 months is by no means large, several conclusions could be drawn from the empirical findings. The VAS was found to deliver quite a number of indicators that contain valid information not only on the economic expectations but also on the heterogeneity of expectations and macroeconomic uncertainty. The rich information set available with VAS overcomes the weaknesses of category-rating scales and can help to extract valuable signals for an early detection of turning points.

As the VAS was found to deliver reliable and valid information on economic expectations, the VAS has been also applied for empirical testing of a set of central assumptions about the data generating process of three-category expectations.

- *Application of the VAS in Research on Expectation Formation*

Chapter 3 reports on an application of the VAS for the testing of a set of assumptions about the data generating process of the three-category measured economic expectations.

The modeling of responses to a category-rating scale implies a latent variable model. One of the main assumptions is the normality of the distribution of respondents' views about changes in the respective variables. Further strong assumptions are made about the indifference thresholds that mark the values of the latent variable at which a respondent is indifferent between two categories. In several methods of quantification of economic expectations it is postulated that (1) the indifference thresholds are symmetrical and (2) neither vary over time (3) nor across the respondents. The study tests whether the normality assumption and the three assumptions about the indifference thresholds hold for business expectations measured by the three-category scale in business tendency surveys.

All above assumptions have been found violated: Business expectations appear to have a systematically varying skewness in different periods of the business cycle. Furthermore, the indifference thresholds of three-category business expectations are asymmetrical, vary across respondents and appear to depend upon time-varying parameters, such as the level of macroeconomic uncertainty and the aggregate level of business confidence.

A particularly interesting finding is that the indifference thresholds are asymmetrical around zero in a way that reflects that respondents weight future losses stronger than future gains. These findings are in line with one of the consequences from the prospect theory of Kahneman und Tversky (1981) that states that responses to losses are more extreme than the responses to gains.

The macroeconomic uncertainty was found to broaden the indifference interval leading to a shift towards the neutral category. These results indicate that the higher the macroeconomic uncertainty the earlier respondents turn to the neutral category

within the three-category scale and they remain within the neutral state until the reference variable (output or profits) overcomes a considerably higher threshold than would be necessary in periods of low macroeconomic uncertainty. Uncertainty also appears to make respondents deflate future gains and losses differently. This empirical evidence may help to explain why peaks are signaled by business surveys with a larger lead than troughs.

The asymmetry between the indifference thresholds of business expectations seems also to be related to the level of the present economic performance. The higher the level of the present business situation, the higher the threshold, which expected changes have to exceed before firms report that they expect an improvement. These results may be explained with the diminishing marginal utility of welfare.

The indifference threshold were also found to vary across respondents. Company size appears rather to narrow the indifference interval, meaning that for larger companies a smaller percentage change in output or profits, results in reporting a fall or a rise on the three-category scale, than for small-scale enterprises.

- *Concluding Remarks*

The empirical evidence of the application of the VAS in surveys on economic expectations is striking. VAS is easy to apply and does not require any quantitative information from the respondents. Delivering the direct measure of business expectations, the VAS supersedes the necessity to make assumptions about the distribution of economic expectations and indifference thresholds, as is required for the three-category based measurement. Furthermore, the VAS was found to be not only a reliable but also a highly efficient measurement instrument of business expectations that delivers a variety of valid economic indicators: The VAS produces valuable information on the dispersion and the skewness of business expectations, providing a comprehensive picture of the businesses' present state and expectations as well as the economic expectations of economists. The VAS dispersion measures also contain information on the heterogeneity of expectations and macroeconomic

uncertainty. Although the time-span is too short to draw conclusions about the forecasting properties of these measures, the results show that VAS indicators explain the rate of change of the industrial production index to a considerable degree, even in the very short time-period.

Also the rapid development of the technological environment offers the VAS a good platform. Research findings indicate that the Internet mode is becoming more and more imperative in business tendency surveys, being a preferable survey mode by a significant proportion of companies (Stangl, 2007).

- *Future Research*

The thesis can be seen as a "roadmap" for further empirical explorations. Many research questions could not be answered due to the still short time-series available at the time the thesis was completed. A future area of research would be to analyze the forecasting properties of the indicators derived with the VAS compared to the traditional category-rating measures.

A highly insightful research area would be to confront the VAS responses with realizations on the company level. At present such data are not available in Germany as business surveys do not collect quantitative information on business realizations.

Furthermore, the thesis demonstrated that the VAS is applicable to the measurement of economic expectations at the company and the expert level. These populations include economists and executives in managerial functions, individuals that are extremely knowledgeable. The application of the VAS for the measurement of economic expectations in the general population or in consumer surveys remains to be tested, to be able to establish the VAS as a generally accepted measurement method of economic expectations.

Chapter 1

Reliability of the Visual Analog Scale for the Measurement of Economic Expectations

Abstract

This chapter introduces the new measurement method of economic expectations – the visual analog scale (VAS) – and evaluates its reliability. Four classes of reliability are examined: Parallel-forms reliability, test-retest reliability, internal consistency and inter-rater reliability. As a side product of the study the reliability of the traditional three-category scale is also evaluated. The data set is based on an international panel of economists and contains 2,470 observations from eight consecutive quarterly survey waves in the years 2005-2007. VAS is found to be a reliable measurement method of economic expectations.

1. Introduction and Background

The role of economic expectations as determinants of economic decisions of consumers, businesses, economic experts and politicians is a matter of particular interest in economic research. Already Keynes (1936) emphasized the role of business expectations in determining output, employment and saving. With Muth (1961) and later Lucas (1972) the concept of rational expectations became one of the central assumptions for many contemporary macroeconomic models. Kahneman and Tversky (1974, 1979) promoted research on bounded rationality and expectation formation under uncertainty, to explain divergences of economic decision making from neo-classical theory.

Because of the prominent role economic expectations are playing in economic theory and in determining economic activity, measurement of economic expectations continues to be subject to discussion. At the individual level, subjective probabilities as a measurement method for economic expectations has moved in the focus of interest. This area of research was initially explored by Juster (1966) and pursued by Dominitz and Manski (1997). At the company and the expert level discrete scales as a measurement method of economic expectations are dominating. The grading procedure here traditionally consists of three general categories: positive replies, indifferent replies and negative replies. The experience of interviews and mail surveys suggested the usage of category-rating scales with three or at most five categories (very good/good, satisfactory, bad/very bad). Although discrete scales imply an information loss, some researchers argue that it is easier and less costly to obtain reliable responses to qualitative questions than to more precise questions (Pesaran and Weale, 2006). There appears to be no dominant argument for the one or the other measurement method of economic expectations and Manski (2004, p. 1369) concludes in his paper on the measurement of economic expectations that people can report their expectations in different ways – as point predictions, verbal assessments of likelihood, or probabilistic expectations.

Visual analog scale (VAS) with anchored ends (bad/worse vs. good/better) has not yet been applied to the measurement of economic expectations, although it is an attractive alternative to category-rating scales, and particularly to the coarse 3-point scale: Respondents can express their expectations on a continuum, which allows for a subtle distinction of their preferences. Furthermore, while the information collected is far more complex than in three-category scales (the values of the VAS saved in the background range from 1 to 100), there is no increase in cognitive load for the respondent (Aitken, 1969). VAS was first described in 1923 by Max Freyd, in his article "The graphic rating scale" in *The Journal of Educational Psychology*. At the present, the VAS is one of the most commonly used measurement instrument of pain intensity in medical research (Jensen et al., 2003).

In the literature on economic expectations and their measurement, the main focus has been placed on validity criteria, particularly rationality and forecasting properties of economic expectations. The key research issue has been to link qualitative expectations data to the real quantitative outcomes. Therefore, the accuracy of expectations has been traditionally addressed by comparing individual or aggregated expectations with realizations (Carlson and Parkin, 1975, Nerlove, 1983, Zarnowitz, 1984, Dominitz, 2001, Manski, 2004). This issue will be discussed in the following chapter of the thesis. The literature on reliability issues and properties of scales measuring economic expectations is rare, if not absent. Also studies working with micro data rather than with aggregated time-series are exceptional. Pesaran and Weale (2006) conclude in their article on survey expectations that the analysis of individual responses to surveys which collect only qualitative information is underdeveloped. This paper adds to the literature on the measurement of economic expectations a comprehensive analysis of scale reliability. The aim of the paper is twofold: The main objective is to explore the reliability of the new scale – the visual analog scale – applied in an economic tendency survey of economists. A side product of the study, however, is a comprehensive test of the reliability of the traditional three-category scale, applied in the majority of business tendency surveys world-wide. Micro-data of eleven consecutive survey waves are used to investigate the reliability of the VAS and the traditional three-category scale.

Reliability refers to the consistency and accuracy of a measurement instrument, and is a necessary condition for validity, as a measurement instrument can not be valid without being reliable. According to the definition, reliability is the correspondence of a value measured by a particular scale, with the hypothetically true value. The main task of a reliability measure is to assess the error component in relation to the true score component. Since the true value is not available, it is impossible to calculate reliability exactly. It has to be estimated in various ways, each representing a different dimension of reliability. There are four general classes of reliability estimates: *Parallel-forms reliability*, *test-retest reliability*, *internal consistency reliability* and *inter-rater reliability*. Reliability of a scale is to be seen as a multi-dimensional construct with each of the above reliability classes exploring a particular dimension of reliability. The present paper investigates all above reliability dimensions.

There are further arguments making this study interesting: (1) Previously the VAS was merely applied in personal interviews in medical settings. In the present study the VAS is applied in a sizeable self-administered Web-survey in a non-medical setting. (2) The study tests an entirely new measurement method of economic expectations, which so far are measured qualitatively (with category-rating scales), quantitatively (by eliciting point estimates) or by subjective probabilities. (3) The data of the study are based on a real panel survey of economists which makes possible to observe long-time effects in the VAS response behavior.

Section 2 of this paper contains a literature review. As literature on the VAS or scale reliability, is almost absent in business tendency surveys research, an excurse is made to the analogical literature in psychological and medical settings. The data used in the study are described in section 3. To fill the lack of literature on reliability of scales measuring economic expectations, section 4 of this paper is dedicated to a discussion of reliability classes and estimation methods applicable to business tendency surveys. Section 5 contains empirical results of reliability estimates. The paper concludes with a discussion of results and inference on whether empirical evidence provides support for the VAS and the three-category scale being regarded as reliable instruments for the measurement of economic expectations.

2. Literature Review

Applications of the VAS in sizeable surveys have emerged recently along with the spread of web-based surveys. Before this survey mode has become popular the use of the VAS in social surveys has been rare due to operational difficulties (Couper et al., 2006). In medical research settings, in contrast, the VAS has been applied for many decades, particularly for the measurement of feelings, like pain and discomfort, and other health state valuations. While search in the ISI Web of Science[1] on the terms "visual analog scale" (or "visual analogue scale") reveals more than 8,000[2] records, a search on the terms "visual analog" (or "visual analogue") and "Internet" reveals only 19 results, indicating that the application of the VAS in Internet-based surveys is a new field of research. Only two of these nineteen studies were applied in non-medical settings (Couper et al., 2006 and van Schaik and Ling, 2003) and only four of these articles dealt with reliability of the VAS as a measurement instrument (Brophy et al, 2004; Coll et al., 2004; Athale et al., 2004 and Lenert, 2000).

Couper et al. (2006) conducted a one-time experiment to explore the utility of the VAS in a Web survey, comparing it to 20-point radio buttons and numeric entry in a text box on a series of bipolar questions eliciting views on genetic versus environmental causes of various behaviors. The authors found the VAS having higher rates of non-completion, higher rates of missing data and longer completion times than the other methods. However, the authors admit that the VAS was tested on a difficult set of items on which respondents are not likely to hold well-formed views: "*The VAS may perform better in situations where the respondent is better able to make fine distinctions among different attitude objects, such as the feeling thermometer ratings of political figures*" (p. 243). In the present study the VAS is applied in exactly this kind of settings: For the measurement of economic confidence of economic experts, a population group that is expected to have well-formed views on the survey topic.

[1] Citation Databases: Science Citation Index Expanded (1995-present), Social Sciences Citation Index (1999-present), Arts & Humanities Citation Index (1999-present).
[2] Couper et al. (2006) reported that the search on "visual analog scale" revealed more than 2,700 citations at the time their paper was written. In the years 2005 and 2006 alone, more than 2,000 scientific studies that applied VAS in their experimental settings have been published.

Van Schaik and Ling (2003) used the VAS and a 7-point Likert scale for the evaluation of human-computer interaction. Participants identified advantages and disadvantages for both response formats. They stated that the fluidity of the VAS response format allowed them a greater range of responses without being restricted to numbers. On the other hand, the VAS was identified as leading to a lack of precision due to the absence of category markers. Problems appeared with the VAS when respondents wanted to give exactly the same response to a different item but could not replicate the precise position of the marker on the VAS. The variability of scores was similar between the two response formats, according to the authors.

The empirical results about the application of the VAS in medical paper-and-pencil surveys are very mixed, according to the three articles containing a critical review of the existing literature on the VAS that have been found by the author. Ahearn (1997) comes to a positive conclusion in her critical literature review on the utility of the VAS for mood measurement. The VAS was found easy to complete and possessing high reliability and validity. Williamson and Hoggart (2005) explored the literature on three commonly used pain ratings scales, the VAS, the verbal rating scale and the numerical rating scale. They also concluded in their review that all three pain-rating scales are valid, reliable and appropriate for use in clinical practice. Butler (1997), in contrast, comes in his critical review on the utility of the VAS for pain measurement to a very critical conclusion.

There are numerous further widely cited[3] studies in which authors concluded that the VAS delivers reliable measures (Gallagher et al., 2002, Bijur et al. 2001., Badia et al., 1999). However critical findings were reported by Wuyts et al. (1999). There are also more recent studies on the VAS reliability in favour (Cook et al., 2001; Happich, 2006; Parkin and Devlin, 2006; Wagner et al., 2007) as well as against the application of the VAS for the measurement of various health states (Svensson, 2000; Brazier et al., 2003).

[3] Cited more than 20 times according to the ISI Web of Science (Citation Databases: Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index) retrieved 01/29/2008.

Although there is a broad literature on the VAS reliability in medical settings, only four articles were found dealing with the VAS reliability in medical Web surveys. They were selected to be discussed in more detail for several reasons: The present study investigates the VAS application in a sizeable Web survey. Web surveys are conceptually different from other survey modes (Couper, 2001), as for example personal interviews, in which the interviewer may help the respondent in the response process. The conclusions of the VAS application in personal interviews may therefore not be transferable to Web surveys. Consequently, from the high variety of literature on the application of the VAS in medical settings those studies will be discussed in more detail which experimental design is equivalent to the experimental design of the present study.

The experiences with the VAS in the four medical Web survey studies were generally positive. Brophy et al. (2004) studied whether the VAS can be used over the Internet for the assessment of disease severity. The Internet version and the paper version were completed twice to assess intra-respondent variation reliability. The authors concluded that assessment of disease severity by VAS may be accurately carried out over the Internet. Coll et al. (2004), reviewed some of the available objective and subjective measures of postoperative pain using the VAS in a Web survey and came to the conclusion that the VAS is methodologically sound, conceptually simple, easy to administer and unobtrusive to the respondent. Also Athale et al. (2004) described a Web-based health self-assessment survey of 43 patients containing VAS for pain, fatigue, and global disease severity, as well as multiple-choice questions from a multi-dimensional health assessment questionnaire. The authors found high reliability scores for pain and global disease severity and moderate reliability scores for the VAS scores for fatigue. Somewhat older study of Lenert (2000) elicited preferences in a research-lab setting with 60 volunteers using the VAS and the standard gamble for the assessment of the subjects' health. For comparison a short health-assessment questionnaire was administered. The VAS was found to be a reliable instrument. Although the results on VAS reliability in these medical Web survey studies were generally positive, it is not clear whether the findings on the utility of the VAS in medical studies can be generalized to social surveys. In medical studies the VAS is applied to the measurement of feelings, in the

case of business tendency surveys, the VAS is applied to the measurement of attitudes towards present and future situations and implies particularly in the latter case a high degree of uncertainty.

There are numerous approaches on the estimation of reliability of the VAS in medical studies, with inter- and intra-rater reliability and the test-retest reliability being the most common reliability forms tested. The majority of studies confirm that the VAS is a reliable measurement instrument, although arguments against the VAS are also present. In the area of business tendency surveys, to the best of our knowledge, there is no literature focusing on reliability criteria of scales at all. Generally, while examining forecasting properties (i.e. criterion validity), reliability is postulated, ignoring the fact that reliability is a basic prerequisite for validity. In this context both research questions – the reliability of the VAS and the reliability of the traditional three-category scale for the measurement of economic expectations – appear to be interesting.

3. Data

The VAS was first implemented in the World Economic Survey (WES) in the April 2005 survey. WES is an international survey of economists which is conducted quarterly by the Ifo Institute for Economic Research. In the present experiment the VAS is applied simultaneously with the three-category scale in two survey questions: *present economic situation* and *economic expectations for the coming six months*. Responses to both questions are used to build an index of economic confidence. This approach sees economic expectations as result of the present economic situation and the respondent's projections. This section describes the two rating scales and the data set, the analysis in this paper is based on.

The software *Stata*^{TM} (Standard Edition 9.2) was used for the entire data analysis, considering $p<0.05$ as significance level.

3.1. The Rating Scales

3.1.1. Three-category Scale (3-Cat.)

Business tendency surveys traditionally apply three-category rating scales.[4] The motivation and popularity of using qualitative rather than quantitative variables in business tendency surveys has several origins (OECD 2003): It is generally much easier for respondents to give qualitative rather than quantitative information; the questionnaire can be completed quicker, as respondents do not necessary need to consult their accounting records. Furthermore linguistic characterizations provide a better mean for phenomena which may be too complex to be described in quantitative values.

However, the three-category rating scales have also several drawbacks: The question format allows the respondents to choose only from the three available options that may not match their real opinions. This leads to a drift towards the central category which includes responses which have positive as well as negative tendency, but which have not reached a particular threshold which would make the respondent to choose one of the extremes. The offering of just three response option results in imprecision and information loss.

3.1.2. Visual Analog Scale (VAS)

Visual analog scales are rating scales on which a subject ranks the preferences along a continuous line or scale. There are numerous variations of the approach (e.g., length of the line, labels for the ends, vertical or horizontal placement, color etc.). These scales correspond to the common usage of measurement scales for physical extensions (thermometer, etc) and are for the respondents easy to understand and to handle. It is a very simple idea, but its benefits compared to three-category rating

[4] See also the ZEW Indicator of Economic Sentiments of the Centre for European Economic Research: http://www.zew.de/en/publikationen/Konjunkturerwartungen/Konjunkturerwartungen.php3

scales are obvious: The VAS enables scores between categories and the respondent can express not only the direction of his attitude but also the magnitude on a 1-to-100 point scale. The scale range allows for a subtle distinction of respondent's preferences. At the same time, by using a graphic scale with few anchors there is no increase in cognitive complexity for the respondent, while the information collected is much broader than any three-category scale would allow.

To emphasize the feature of symmetry the scale received three anchors a dotted line characterizing the middle and verbal descriptions of the two extremes (see Figure 1.1). The software was written to record slider movement on a response continuum of a 1-to-100 point scale. While the practical application of such scales was difficult in mail and fax surveys, due to high operational costs, the Web provides a well-suited environment for the VAS implementation in a sizeable survey.

Taking into account the experiences of previous Web studies[5], a version of the VAS was chosen where the marker is set by clicking on a desired position on the slider. By clicking on another position the marker jumps to a new position and by a double click it disappears.

One of the necessary prerequisites for an accurate measurement is that respondents are motivated to use the measurement instrument. In the literature respondents' motivation has been merely addressed through the measurement of completion time needed to respond to a particular scale (Couper et al., 2006). However, longer completion time may not be regarded as a burden and negatively affect respondent's motivation if the question appears interesting to the respondent. In contrast, respondents may prefer to take more time to complete the question. Therefore, in the present experimental design the completion time has not been measured. However, a *comment field* below the VAS was offered, where respondents could input any kind of comments. Furthermore, in the background, not visible to the respondent *number of clicks* has been registered, which can be used as an indicator of how fast the respondents learned how to use the instrument and after repeated use, how fast they

[5] Van Schaik and Ling (2003) reported about respondents having difficulties with the use of a slider, where respondents had to move the marker, by holding the left mouse button and simultaneously move the mouse. This handling problem was avoided by applying the above described VAS version.

formed an assessment on the VAS. *Item non-response* can also be used as an indicator for the respondents' motivation. However, in the present study the VAS and the three-category scale have been applied simultaneously on the same questionnaire. In the introductory note it was also explained to respondents that the VAS is an experiment for testing a new measurement method. Consequently, non-response to the VAS could not be disentangled from non-participation in a voluntary experiment and is thus not used as an indicator of respondent's motivation in the present study.

There were two different types of non-response to the VAS. The first kind of non-response appeared, when respondents initially responded, but then reversed their intention and deleted their response, which rarely occurred. In all eight survey rounds in only 27 of 2,470 questionnaires (less than 0.2%) the initial response on the analog scale has been deleted by the respondent. The second kind of item non-response occurs when respondents do not respond to a questionnaire item at all. This non-response can have various reasons, not necessarily related to the VAS as a measurement instrument. The analog scale was positioned in both surveys at the very end of the questionnaire, being the last question of the survey. Furthermore, both VAS questions were redundant, to the questions asked on the traditional three-category scale at the very top of the questionnaire. And, participants were informed in the introductory notes that the two VAS questions are part of an experimental study. Consequently, some respondents were reluctant to answer to the same question for a second time, although the introductory note briefly explained the research aim. Thus, this type of non-response may rather be classified as non-participation in the experiment. However, there were also pure technical reasons for non-response to the VAS. Some respondents had particular Internet settings installed at their PC which evoked error messages relating the VAS. In the course of the experiment these technical problems have been solved and the non-response to the analog scale slipped below 5% (in the third and fourth quarter 2006). However, another technical problem occurred, as the new version of the Firefox Internet browser that become popular in 2006 blocked the scale and non-response jumped back to 9% in the first quarter 2007 (see Annex, Table 1.1). Also this technical

problem has been solved and in the second quarter 2007 non-response to the VAS declined again.

The comments retrieved in the comment field provided below the VAS indicated that respondents found the VAS "self-explanatory" and "interesting". All comments received during the eight quarters of the survey have been classified in positive relating the scale (such as "interesting method", "good idea", etc), neutral (relating other aspects of the questionnaire or the country's economy) and negative (such as "old method better" etc.). Table 1.2 in the Annex summarizes the frequency of positive, neutral and negative comments to the new scale, over the eight survey waves. Most comments on the VAS were received in the first two survey rounds and were with one exception positive. In the following survey waves the comment field was used to express other information rather than the opinion on VAS or their experience with the VAS. This result indicates that repeated application of the VAS in a panel survey did not elicit negative experience. There were seven comments which indicated that respondents had problems with the VAS. Three respondents thought that they were supposed to move the marker, which did not work as they had to click it according to the instruction in the question wordings. Four further respondents indicated that the VAS was defect. The reason was that a new version of the Firefox browser caused incompatibility errors in the VAS software. In the following surveys this problem was eliminated. Table 1.3 in the annex lists all comments received during the eight survey waves.

In the background, not only the responses to the VAS were saved, but also the number of clicks each respondent did, were recorded. In the first survey round, in the second quarter 2005, the respondents were confronted with the new scale for the very first time. At this point of time, the interpretation of the number of clicks is ambivalent. On the one hand, a low number of clicks would suggest that a respondent learned fast how to use the instrument and formed the assessment on the VAS intuitively. On the other hand, a high number of clicks in the very first survey round may also indicate interest in the new scale. For both situations, one would expect the number of clicks to decrease in the course of the study as participants become familiar with the new instrument. Table 1.4 in the Annex summarizes the

average and the maximum number of clicks to the two questions over the course of the study. The number of clicks varied in the very first survey round between 1 and 18 clicks for the variable present economic situation and between 1 and 10 for the variable "economic expectations". The data show that the number of clicks indeed decreased over time and stabilized after the third round at around two clicks for the variable "present economic situation". For the question on economic expectations the number of clicks remained more or less stable at around five clicks, on average. These results show that the judgment and expectations variables elicit different response behavior. The present economic state, that is known by the respondents, can be reflected very fast, with one or two clicks, on average. The future economic situation is, however, uncertain. The higher number of clicks in this variable may reflect this uncertainty. This issue will be discussed in more detail in the next chapter of the thesis.

3.2. Ifo World Economic Survey (WES)[6]

The Ifo World Economic Survey (WES) has been conducted since 1983. The survey assesses worldwide economic trends by polling national and transnational organizations worldwide about current economic developments in their respective country. Since 2002, the WES panel stabilized at about 1,000 economists from more than 90 countries responding quarterly. WES is a mixed-mode survey with circa 65% of the panel participating via mail or fax, circa 5% via E-mail and ca 30% via Web (see Annex Figure 1.1). The data set of the present study contains only responses to the Web questionnaire, as the VAS was implemented only in the Web form.

The WES questionnaire asks mainly for qualitative information, as, for example, for the assessment of the country's general economic situation and expectations regarding important economic indicators. The WES panel contains economic experts with a range of specializations in management, finance, and other business functions. About 65 percent of the WES panelists work for international corporations – companies (circa 45%), banks (circa 15%) and insurance (circa 5%). Some work in

[6] Stangl (2007) contains a detailed description of the WES micro data.

economic research institutes (circa 10%) and chambers of commerce (circa 10%), consulates and embassies (circa 5%). The remaining 10% are affiliated with international organizations (OECD, IMF, Asian Development Bank etc.), foundations, media or small scale enterprises. Although the panel members are heterogeneous with respect to their professional affiliation, all respondents are highly qualified, being in a leading position or occupied with economic research within their institution. The survey participation is absolutely voluntary and derives entirely from the interest in the survey topic.

The VAS has been introduced in the April 2005 Internet survey round. The data set of the present study contains 2,470 observations from eight consecutive quarterly survey waves: Q2/2005, Q3/2005, Q4/2005, Q1/2006, Q2/2006, Q3/2006, Q4/2006 and Q1/2007, as well as data from the re-test reliability experiment implemented in the Q4/2007 survey round with 342 observations. The number of responses each quarter ranges from 280 responses, received in the Q3/2005 and 326 responses received in the Q1/2006.

4. Reliability Estimation Methods

4.1. Parallel-forms reliability

Parallel-forms reliability is used to assess the consistency of results from two scales measuring the same content. Both variables of interest – assessment of the present economic situation and economic expectations – are measured simultaneously by two methods: the three-category scale and the VAS. The two scales have been compared with each other with respect to *consistency, degree of agreement* and *means over time*. The first test examines whether respondents respond to the two questions consistently. The second test uses Spearman's rho to estimate the correlation of responses elicited by the two scales. Spearman's rho is an ordinal measure of association and has been chosen throughout the study for several reasons:

First, it is the most common correlation coefficient for two ordinal variables or an ordinal and an interval variable. There is little agreement in the literature as to whether the VAS ratings have ordinal-level, interval-level or even ratio-level scores, because the VAS responses can be converted to a continuum of numerical values (as a rule on a 1-to-100 scale). Butler (1997) concludes in his critical review of the VAS application in pain assessment that the status of analog scaling appears to be somewhere between an ordinal and an interval measurement. Rank-based measures of associations, like Spearman's rho, would be in either case the most appropriate statistics. In the third test the *means are compared over time.* High parallel-form reliability coefficients would indicate that the two scales are very similar, are measuring the same and could be used interchangeably. However, this kind of reliability test can only establish evidence on whether two scales are equivalent. Given the case that the results from the two scales fall apart, the parallel-form reliability tests are incapable of establishing any prove on which scale is more reliable.

4.2. Test-retest Reliability

Test-retest reliability is an important way for the assessment of a measurement instrument. This reliability dimension receives particularly much attention in the literature on the VAS (Aitken, 1969, DeLoach et al, 1998, Bijur et al., 2001), as it is expected that it is more difficult to replicate a response on a VAS than on a category-rating scale. These studies attempt to compute an imprecision interval and a minimum significant difference of VAS scores. Test-retest reliability is based on the elicitation of a response to a scale across different times and refers to the degree with which the repeated measurement yields the same result. This approach assumes that there is no substantial change between the two occasions – a reliable measurement instrument will elicit stable responses only given the case that the value of interest remained stable. Enterprises are highly dynamic entities and the change of the economic situation of an enterprise (additional orders, investment decision etc.) may appear at any time. The same applies to the assessment of the economic situation of a country. The surveyed expert may receive new information immediately after the

completion of the questionnaire and up-date his views. Consequently, in the present study, respondents were asked immediately after the completion of the questionnaire to reproduce their response on the VAS. There are numerous ways to estimate the test-retest reliability statistically. The correlation coefficient between the two sets of responses and intra-class correlation (ICC) are often used as a quantitative measure of the test-retest reliability. Further quantification method used to establish the test-retest reliability is the root of the squared difference between the initial response and the re-test response. An important question in this context is whether reproductions appear to be to the same degree accurate at different areas of the scale, depending on the distance to the three anchors (middle line and the ends). Further explanatory variables of the test-retest reliability are also investigated.

4.3. Internal Consistency

Internal consistency is a further common way to assess reliability, which is to be seen as complementary to the above described estimations. Internal consistency estimates are usually based on the correlation among different items of the same construct. The questionnaire contains two general types of variables – assessments and expectations. The first type of variables aims at assessing the present economic situation, the later future economic conditions. In the Ifo World Economic Survey economic experts are not only asked to assess the present and the future economic situation in general, but also to rate various aspects of economy, such as capital expenditures, private consumption, foreign trade, monetary policy and political climate. The variables "overall economic situation" and "overall economic expectations" summarize these dimensions and are consequently expected to be correlated with the different items, measuring various dimensions of economy. In the first step, *Spearman's rho* correlations between the two variables of interest (present economic situation and economic expectations) measured by the two scales (the VAS and the three-category scale) and the other items representing various dimensions of economy are calculated. In the second step the items representing one construct (either assessment of the economic situation or economic expectations) are analyzed whether they contain a single latent variable – a common factor. To estimate the extent to which

the set of items measures a single unidimensional latent variable *Cronbach's alpha* is used. *Cronbach's alpha* will generally increase up to 1 when the correlations between items increase. As a rule of thumb (Peterson, 1994), an *alpha* of 0.70 or higher is required to establish evidence for a unidimensional construct. Resulting from the *alpha* analysis one common factor variable has been extracted from the set of items representing different dimensions of the economic situation in a country. In the third step, the relationship between the present economic situation measured by the two scales and the common factor variable was investigated.

4.4. Inter-rater reliability

Inter-rater reliability or inter-rater agreement establishes the extent of how much agreement or consensus there is in the ratings given by judges on the same subject. The data of the present study are from a survey that is conducted in the same methodology in more than 90 countries and polls several experts per country. As the surveyed economists assess the economy of the same country, they, theoretically, given the same ability and knowledge, should come to a high degree of agreement. Responses across countries, in contrast, are expected to diverge, in case the economic situation differs between the countries. [7]

For each scale there are at least two measures of agreement that can be calculated: For three-category data, consensus can be measured as number of agreements divided by total number of observations or intra-class correlation (ICC). For the interval-based VAS, consensus can be measured by the Pearson correlation coefficient or ICC. The ANOVA-based ICC coefficient measures the ratio of between-groups variance to total variance. The formula used for the ICC (Shrout und Fleiss, 1979) is: $\sigma^2(b)/[\sigma^2(b)+ \sigma^2(w)]$, where $\sigma^2(b)$ is the variance of the trait between groups (countries) and $\sigma^2(w)$ is the variance within groups. The ICC will increase up to 1 when there is little variation within a group, e.g. if all raters within a country

[7] There is a broad literature on disagreement in economic expectations and reasons of disagreement. Relying on Doepke and Fritsche (2006) arguments, the following assumptions are made: the economists have same ability and knowledge about the current state of the economy, rely on the same assumptions on the path of exogenous variables and apply the same model of the economy. Furthermore, it is assumed that there is variation in business conditions across countries.

give the same, or similar scores. The ICC has been chosen for two reasons: First, we are interested in comparing the VAS with the three-category scale and the ICC can be used to compare the reliability of different instruments. Second, the degree of agreement and Pearson correlation can be only applied with two raters. An alternative would be to calculate the correlations between all pairs of raters. The striking difference between the ICC and the correlation coefficient, however, is that the latter method ignores the between country variance. Furthermore, as only eight survey waves are disposable at the time the paper is written, and not all respondents participated regularly, the data set would contain too few observations per respondent. Consequently, ICC appears to be the most appropriate statistical method for estimating the inter-rater reliability in the present data set. Although an ANOVA based method is suitable for interval-scaled data, it was also applied to responses derived with the three-category scale, under the assumption that the data generating process of the three-category responses is based on a latent continuous variable.

According to the Shrout und Fleiss (1979) classification of ICC for reliability, the unique rater design is applied, i.e. each country is rated by a different set of respondents. The ICC was computed by a random effects ANOVA model, using the between country variance and within country variance (measure of the error variance). Ideally the within country variance is expected to be independent of the variables absolute value, i.e. the error magnitude is expected to be independent from the level of economic sentiments (both present and expected). Consequently, before calculating the ICC the VAS data were tested for heteroscedasticity.

5. Results

5.1. Parallel-forms Reliability

5.1.1. Inconsistent Responses to the VAS and the Three-category Scale

The first parallel-forms reliability test estimates whether the two scales measure the same variable consistently. Inconsistencies are contradicting responses and can occur when a positive assessment is given on the three-category scale and a negative on the VAS or vice versa. There may be various reasons for inconsistencies within the questionnaire. The most common reasons are mistakes and attention deficits on the respondent side.

Table 1.5 in the Annex summarizes the frequency of inconsistent responses. In the variable present economic situation inconsistent responses are relatively rare (as a rule less than 2%). For, the variable economic expectations inconsistent responses occurred more often, in up to four percent of responses. The higher frequency of inconsistent responses in the variable on economic expectations may again be an indication of uncertainty or respondent's ambivalence relating future economic conditions. For the following analysis questionnaires with those inconsistent responses have been removed from the data set.

5.1.2. Degree of Agreement Between the VAS and the Three-category Scale

The correlation analysis investigates whether a higher assessment on the VAS corresponds to a higher assessment on the three-category scale and vice versa. For the whole data set the correlation between the three-category and the VAS for the variable present economic situation is with 0.8 (N=2,158) and for the variable economic expectations with 0.7 (N=2,151) very high. Also within the countries (see Annex, Table 1.6), which are in WES the sample units, the correlations are positive,

in the majority of cases significant and relatively high (generally over 0.5). Although, the Spearman's rho is the most appropriate correlation measure for the estimation of the degree of agreement, the high number of ties in the three-category scale makes the use of it problematic. Consequently, the parallel-form reliability is also estimated looking at the aggregate time-trends.

5.1.3. Comparison of the VAS and the Three-category Scale Results Over Time

The Figure 1.3 in the Annex illustrates the means of the variables present economic situation and economic expectations over time, for the main OECD countries. The figures show a high correspondence of the economic confidence measures derived by the two scales over time. This relationship holds particularly if the number of respondents is sufficiently high. However, in some countries the indicators derived with the three-category scale and the VAS diverge, as for example in Ireland. In Ireland all respondents stated throughout the experimental period, that the present economic situation is good and will remain stable. Consequently there was no variety in the index derived with the three-category scale, while the index derived with the VAS fluctuated, coming closer to the changes in the real GDP growth in Ireland (see Annex, Figure 1.4).

In summary, the results of the parallel-reliability tests indicate that the responses elicited with the VAS and the three-category scale are highly correlated. This positive relationship is stable over time and is true across countries. However, further reliability estimates are necessary to establish evidence, whether the scales measure the variables in a precise way.

5.2 Test-retest Reliability of the VAS

5.2.1. Experimental Settings

The test-retest reliability was estimated by the following experiment: In the 11th survey round (Q4/2007) the respondents were asked to replicate their response on the VAS. After respondents have answered the standard questionnaire containing the two VAS questions and pressed the "send" button, a new website appeared with an explanatory note and the request to participate in a one-time experiment. The respondents were asked to repeat their assessment on the VAS without switching back to the standard questionnaire in their browser, in order to view their initial response.[8] Figure 1.5 shows a screenshot of the re-test experiment.

In the 11th survey round, in October 2007, 342 responses to the standard questionnaire have been collected. Almost 80% of respondents (269 respondents) also voluntary participated in the re-test experiment. A t-test was performed to test whether the respondents differed from non-respondents. The results are summarized in the Annex, Table 1.7. Except for two variables, the responses of non-participants in the re-test experiment do not significantly differ from the responses of participants. These results justify the assumption that non-participation is random.

The data have been also proved for inconsistent responses of the type that a positive response has been given to the VAS question on the standard questionnaire and a negative response to the VAS question in the re-test or vice versa. There were only two respondents who gave inconsistent responses in the re-test experiment. The distance between the two responses was however relatively small. These responses were consequently not regarded as inconsistent.

[8] If respondents ignored this request and switched to the previous page, they were excluded from the experiment and were forwarded to the final standard page where they usually can change their contact details and are instructed to exit the browser.

An interesting finding was that in the re-test experiment, respondents needed a lower number of clicks to form their response on the VAS than in the standard questionnaire. These findings indicate that respondents, on average, exert a greater effort to identify the initial marker position on the slider that corresponds best to their economic confidence. A replication of their marker position costs them less effort.

5.2.2. The Estimation of the Imprecision Interval of the VAS

The means of the two VAS questions – "present economic situation" and "economic expectations" – in the standard questionnaire and in the re-test experiment did not significantly differ (see Annex, Table 1.8). These results indicate that when the responses on the VAS are reproduced the error variance does not bias the aggregated results.

The distance between the initial VAS response in the standard questionnaire and the VAS response in the re-test experiment can be viewed as the imprecision interval of the VAS and is calculated as the absolute difference between the two scores. Table 1.9 in the Annex shows the distribution of the VAS imprecision interval. As it was easier for those respondents to meet the anchor position in the re-test experiment whose VAS response in the standard questionnaire was on one of the three anchors or close to them, the group of respondents whose marker position was farther away from one of the anchors is displayed separately. A relatively high proportion of all participants (25.4 percent in the variable "present economic situation" and 33 percent in the variable "economic expectations") met exactly the same marker position in the initial VAS question and in the re-test VAS question. However, in 18 percent of responses (48 responses) in the variable "present economic situation" and in 31 percent of responses (81 responses) in the variable "economic expectations" the marker position on the slider was placed at one of the three anchors (the middle anchor or one of the two extremes) or close to it (+/-1 point of the scale). Among respondents whose marker position in the initial VAS question was not on one of the anchors, the proportion of respondents who met exactly the same marker position was also considerably high (18.5% in the variable "present economic situation" and

22.4% in the variable “economic expectations). However, there were also six outliers, i.e. cases in which the first response differed from the re-test response by more than 10 points on a 100-point scale.

Allover, more than 50% of the paired measurements were within 1 point of one another, circa 90% were within 4 points, and circa 95% were within 6 points. These results appear comparable to the results of similar medical studies. Bijur et al. (2001) estimated the reliability of paired VAS measurements obtained 1 minute apart from a convenience sample of adults with acute pain. At a 100mm VAS, 50% of the paired measurements were within 2 mm of one another, 90% were within 9 mm, and 95% were within 16 mm. The authors concluded that the VAS is sufficiently reliable to be used to assess acute pain. DeLoach et al (1998) found lower re-test reliability with 30% of the paired measurements within 5mm and 92% were within 20mm in a sample of patients with acute postoperative pain. Wagner et al. (2007) in contrast found a very small mean difference in the VAS score between tests of less than 1mm in their study of mountain sickness.

The mean difference between the initial VAS response and the re-test response is in the present experiment 1.8 for the variable “economic situation” and 1.6 for the variable “economic expectations” (see Table 1.10). Although the absolute difference in both questions is significantly different from zero, this imprecision interval is relatively small given a 100-point scale. As expected, the imprecision interval of those respondents whose initial VAS response was on an anchor or very close to it is significantly smaller, than the imprecision interval of respondents whose marker position was farther away from one of the three anchors. In fact, the imprecision interval is not significantly different from zero in the group of respondents whose initial VAS response on the standard questionnaire was close to an anchor (0.56 for the variable “economic situation” and 0.49 for “economic expectations”).

5.2.3. Factors Explaining the Imprecision Interval of the VAS

In the following test it has been investigated which further factors are related to the width of the imprecision interval in the re-test experiment. Allover, six variables were included in the analysis. In the preceding analysis distance to an anchor and position of the marker exactly on one of the anchors were found to influence the width of the imprecision interval. The variable "distance to an anchor" ranged from zero to a maximum value of 25 (given the middle anchor at point zero and the extremes at -50 and +50). The variable "marker position at the middle anchor" was included as a dummy. Further variables were: frequency of participation in the survey during preceding eleven survey rounds with VAS questions in the standard questionnaire, dummy variable identifying whether respondent had to be reminded (1) or responded timely (0), the imprecision interval in the other variable and the difference to the marker position of the other variable. The reason for including the last variable in the model was that the marker position of the one variable could serve as an additional anchor, as the two VAS variables were presented below each other in the questionnaire. The marker in the first variable may serve as an anchor particularly if the markers are in the same area of the scale. The imprecision interval of the respondent in the other variable was included to test whether the imprecision interval is correlated within respondents. The model was estimated by linear OLS. The results are summarized in the Annex, Table 1.11.

As expected, the higher the distance of the marker position from one of the anchors in the initial VAS question of the standard questionnaire, the higher the imprecision interval. Also those respondents who had the marker at the middle anchor or close to it (+/-1 points) had a lower imprecision interval.

Respondent's experience with the survey is one of those factors that were found to be significantly related to the imprecision interval in the variable "present economic situation". The more often a respondent had participated in the survey during the experimental period of 11 survey rounds in which the VAS was included in the standard questionnaire, the smaller was the imprecision interval on the VAS. This

result is encouraging for panel-based surveys as it indicates that reliability of the VAS increases in panel settings.

It has also been investigated whether respondents who had to be reminded and thus may be less motivated to respond to the questionnaire in general and VAS experiment in particular, have a higher imprecision interval. However, the fact whether a respondent was in the group of participants who had to be reminded, had no significant effect on the respondent's imprecision interval.

Respondents who had a smaller imprecision interval in the variable "economic situation" had a smaller imprecision interval in the variable on "economic expectations" and vice versa, holding other factors constant. These results indicate that the width of the imprecision interval varies stronger across individuals than within individuals and may depend upon unobserved personal characteristics, as for example age, participant's computer skills or some external factors such as the quality of the display or the mouse. This question is worth deeper investigations in future research on the VAS application in social surveys.

The difference between the marker position in the variable "economic situation" and the marker position in variable "economic expectations" had no significant effect on the imprecision interval.

The factors included in the above model explain about 11% of the imprecision interval variation in the variable "economic situation" and about 12% of the imprecision interval variation in the variable "economic expectations".

The overall results indicate that the VAS has a high re-test reliability, with an impressively small imprecision interval, which is comparable or even smaller than the estimates from medical studies and which declines in panel settings.

5.3. Internal Consistency

5.3.1. Pairwise Correlations of the VAS and the Three-category Scale with other Variables Measuring the Same Theoretical Constructs

In the first step, the variables present economic situation and economic expectations measured by the three-category scale and the VAS were correlated with other judgment and expectations variables measuring different areas of these same two constructs. Table 1.12 in the Annex contains the Spearman's rho pairwise correlations between the variables "present economic situation" and "economic expectations" measured by the three-category scale and the VAS and the other judgment and expectations variables from the standard questionnaire. As expected the judgment variables are more strongly correlated with the assessment of the present economic situation than with economic expectations and vice versa. The differences in correlations of the three-category based measures and the VAS based measures with the various dimensions of the present and future economic conditions are relatively small. In the next step Cronbach's alpha was used to assess whether the various dimensions of the present economic situation and economic expectations can be combined to a single common factor. The term "common factor" is here refering to a latent variable that is measured by a set of items. The latent variable that is measured by the set of judgment variables listed in Table 1.12 is the "overall economic situation". The second latent variable is the "overall economic situation in the next six months". Cronbach's alpha estimates the extent to which a set of items measures the latent variable. Technically speaking Cronbach's alpha is the average of inter-correlations among the items. As a rule of thumb, an alpha of at least 0.7 or higher is required to establish indicator reliability (Peterson, 1994). The results of the Cronbach's alpha analysis are summarized in the Annex, Table 1.13. Only the common factor "overall economic situation" shows a sufficiently high alpha of 0.78. The alpha of the "overall economic expectations" capturing only four variables is with 0.68 not sufficiently high to be used as a common factor. As result, only one variable representing the common factor of the "overall economic situation" has been constructed as the mean of the standardized values of seven variables from Table

1.13. In the next step it is investigated, how strong the variable present economic situation, measured by the three-category scale and the VAS, correlates with this generated common factor variable.

5.3.2. Correlation of the Generated Common Factor with the Variable "Present Economic Situation" Measured by the VAS and the Three-category Scale

Table 1.14 contains two measures of association between the generated common factor variable (*overall economic situation*) and the variable "present economic situation" measured by the two scales (three-category scale and VAS). The Spearman's rho was again used as a rank based measure of correlation. However, as the common factor variable comes close to an interval scale, the table contains also R^2 that was calculated from the regressions of the common factor variable on the variable "present economic situation" measured by the three-category scale and the VAS. R^2 estimates the share of variance of the common factor variable that is explained by the three-category scale variable or the VAS. The results are summarized in the Annex, Table 1.14. Both measures of association with the common factor variable are in almost all quarters, except Q3/2006 and Q1/2007, higher if the variable is measured by the VAS. Although, the difference between the measures of association of the common factor variable with the three-category scale and the VAS based measurement of economic situation is not significant, on average. However, all variables on the questionnaire are measured by three-category scales and the correlation between them may be enforced by anchoring effects resulting in higher correlation with the three-category based measure of economic situation. Nevertheless, the VAS has demonstrated a high internal consistency reliability which is not inferior to the traditional measurement.

5.4. Inter-rater Reliability

In the frame of the survey a group of economists is surveyed in each country. The "*between country variance*" is the natural variation between the countries. The

"*within country variance*" represents the error variance of the measurement, a measure of disagreement between respondents within a country. The inter-rater reliability measure ICC (intra-class correlation coefficient) is a measure of the ratio of the *"between country variance"* and the total variance, including the *"within country variance"*. The greater the ratio between these two components, the higher is the ICC, which can take values between zero and one. The ICC can be computed by a random effects ANOVA.

As the ICC coefficient is responsive to outliers with regard to the difference between the individual value and the country mean, observations with individual's value deviating very strongly from the country mean – far above the limit of the upper (99^{th}) and the lower (1^{st}) percentile – have been removed from the data set. However, there were only few outliers in the dataset. In total only 22 of 2,341 observations were dropped. Only countries with more than five respondents were considered in the following analysis, in order to have a comparable measure of the within country variance for different countries. The data set for the ICC analysis contained 1,319 responses from 26 countries and is summarized in Table 1.15 of the Annex.

Ideally the within country mean square variance is expected to be independent of the variables absolute value. To test for heteroscedasticity the data have been plotted in a scatter plot showing the deviation of the individual value from the quarterly country mean as a function of the quarterly country mean. Independent of the country's mean level of economic confidence, the within country error variance appears to be random (see Annex, Figure 1.6).

Table 1.16 in the Annex contains the detailed ANOVA results and the formula for the computation of the ICC. The resulted ICC are summarized in Table 1.17. The ICC coefficients are somewhat higher for the variable "present economic situation" than for the variable "economic expectations". This can be explained by the fact, that ICC decreases as the "*between country variance*" decreases and the "*within country mean square variance*" remaining stable. The variation of economic expectations across countries is generally lower than the variation of the countries' present economic states across countries (see Table 1.16). This difference may be explained

by the fact that economic expectations are more strongly influenced by global economic sentiments than the countries' present economic states.

The ICC measures are similarly high for both scales (Table 1.17). The two scales do not significantly differ, on average, relating the inter-rater reliability (according to a two-tailed t-test). The inter-rater reliability results indicate that despite a higher range of response options, the VAS has a high power in measuring the differences of economic confidence across countries. However, also the traditional three-category scale is capable in distinguishing across countries, despite its limited range.

6. Discussion

Respondents' motivation to use the VAS was high, as reflected by numerous positive feed-backs in the comment field and participation in the VAS experiment, although the positive experience applies to a canny application of the VAS in the survey. There is a possibility that respondent compliance may be compromised with multiple application of the VAS (VAS batteries), as has been experienced by Couper et al. (2006).

The number of clicks suggests that respondents form the assessment on the VAS fast and intuitively. A higher number of clicks in the variable "economic expectations" reflects uncertainty about the future economic conditions rather than difficulties with the analog scale and can be used in future research to derive additional uncertainty measures on the variable of interest.

The results from the *parallel-forms reliability* analysis indicate that responses elicited by the two scales are highly correlated. The analysis of the resulted time-series reveals that the indexes derived by the three-category scale and the VAS are highly correlated too. As this kind of reliability estimates can only establish evidence on how strong two scales are corresponding to each other, but not respond to the question, whether the scales contain a high error component in its measures, three further reliability analyses have been undertaken.

The *test-retest reliability* analysis demonstrated that the VAS has an impressively small average imprecision interval of less than 2 points on a 100-point scale (more 50% of the paired measurements were within 1 point of one another, circa 90% were within 4 points, and circa 95% were within 6 points). This imprecision interval does not bias the aggregated results. The width of the imprecision interval decreases in panel settings with respondents participating more often in the survey having a smaller imprecision interval than irregular participants. In areas of the VAS which were close to an anchor the imprecision interval was even not significantly different from zero. The VAS test-retest reliability results correspond or are even better than the findings in similar medical studies (Bijur et al. 2001, DeLoach et al, 1998, Wagner et al, 2007).

Both scales were found to deliver *internally consistent* measures of the present economic situation and economic expectations. Comparing the correlation coefficients between the various questionnaire items and the variable "present economic situation" measured by the three-category scale and the VAS, for the majority of items the Spearman's rho was higher when the variables were measured by the VAS. From the set of items measuring various dimensions of the common factor "overall economic situation" a common factor indicator was generated applying Cronbach's alpha. Both scales deliver measures of the variable "present economic situation" that explain the common factor variable to a sufficiently high degree.

The results of *inter-rater reliability* analysis using intra-class correlation coefficient (ICC) revealed that the two scales do not significantly differ relating the inter-rater reliability. These results show that despite the higher range of response options available with the VAS and its superior ability to capture subtle differences across responses, the inter-rater agreement does not substantially decrease.

Summarizing the results, the VAS was found to be reliable measurement instrument of the present economic situation and economic expectations. However, reliability refers to the consistency and accuracy of a measurement instrument, and does not imply validity. The present study established the reliability of the VAS as a

measurement instrument of economic expectations and provided a sound basis for further research on the ability of the VAS to measure economic expectations. The next step in examining its potential is the analysis of validity of indicators derived with the VAS. Validity describes how well an indicator measures what it is supposed to measure. This area of research encompasses a number of tests in which VAS indicators is confronted with reference criteria and is covered in the next chapter of the thesis.

Annex Chapter 1

Figure 1.1. Extract from the WES Web Questionnaire: Three-category Scale and the VAS Applied to the Questions on the Present Economic Situation and Economic Expectations

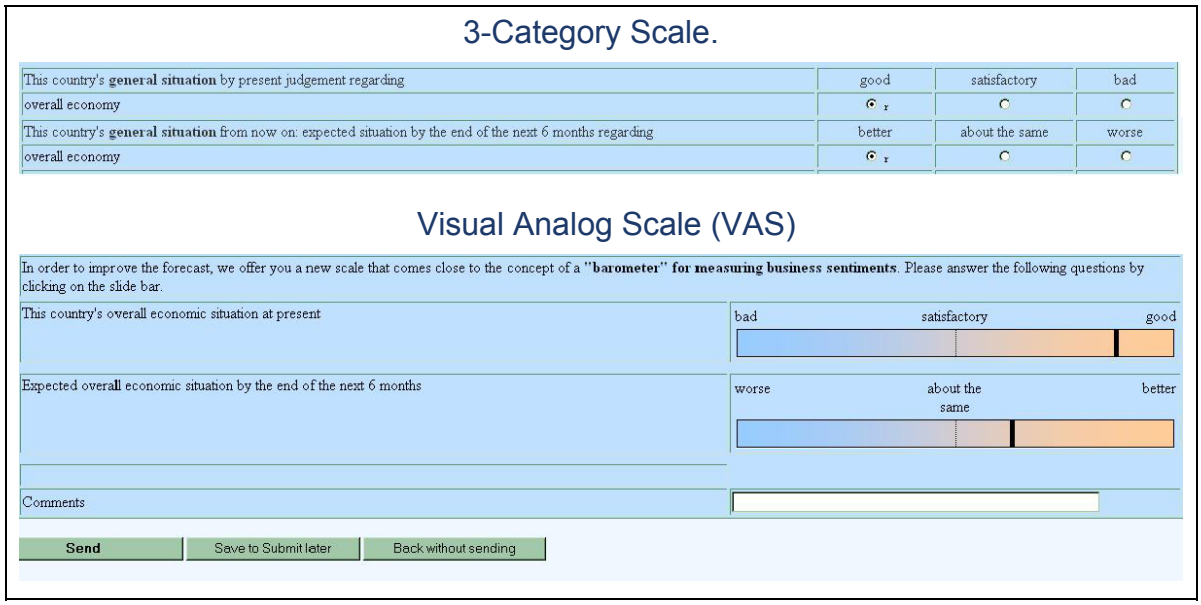


Figure 1.2. Proportion of Responses to the WES Modes: Web, Mail/Fax and E-mail

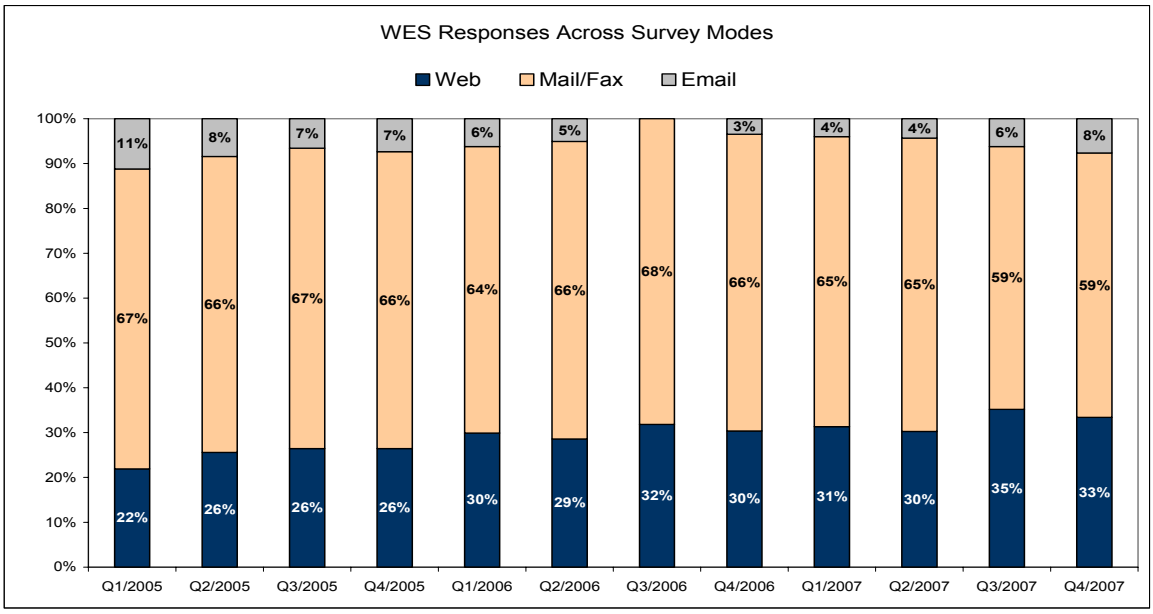


Table 1.1. Non-response to the Three-category Scale and the VAS in the Variables "Present Economic Situation" and "Economic Expectations"

		Three-category Scale				Visual Analog Scale			
		Present		*Expectations*		*Present*		*Expectations*	
Quarter	**Obs.**	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*
Q2/2005	289	1.0%	3	1.4%	4	12.5%	36	12.5%	36
Q3/2005	280	1.1%	3	1.1%	3	9.3%	26	9.6%	27
Q4/2005	293	0.0%	0	0.0%	0	7.8%	23	8.5%	25
Q1/2006	326	0.0%	0	0.9%	3	6.1%	20	6.4%	21
Q2/2006	315	0.0%	0	0.3%	1	6.7%	21	7.0%	22
Q3/2006	321	0.3%	1	0.6%	2	5.0%	16	4.4%	14
Q4/2006	322	0.0%	0	0.0%	0	4.0%	13	4.3%	14
Q1/2007	324	0.3%	1	0.3%	1	9.0%	29	8.6%	28

Obs. = Number of observations

Table 1.2. The Frequency of Positive, Neutral and Negative Comments, as well as Comments on Problems with the VAS

Quarter	**Positive comments**	**Neutral comments**	**Negative comments**	**Problems with VAS**
Q2/2005	13	29		1
Q3/2005	7	24	1	2
Q4/2005	2	13		1
Q1/2006	3	20		
Q2/2006	1	19		
Q3/2006		25		
Q4/2006	1	14		1
Q1/2007	1	23		2
Total	**28**	**167**	**1**	**7**

Table 1.3. All Comments Received During Eight Survey Waves

Positive Comments

"A very good improvement."
"Better system."
"Congratulations for the new barometer experiment."
"Good idea if you can quantify."
"I think the new scale is a good solution."
"In my opinion the scale above is very good for measuring business."
"In our opinion the new scale would be a much better way of showing single situations."
"Interesting idea."
"It is good way to measure the business sentiments by barometer but please add the numerical scale."
"Nice barometer."
"Nice new scale."
"The barometer feature is useful because it allows the respondent to give a more accurate response."
"The scale-method is worth testing."
"Barometer is far more precise."
"Good measure."
"I appreciated the new visualizing approach in Question 11."
"I hope this will work, looks promising."
"I like the barometer. It is a good idea to introduce it."
"In my opinion the barometer is better and so easy to see the results."
"The barometer is interesting. It might be better a numerical scale at the bottom 0-100%."
"Measuring business sentiments is better."
"Nice."
"Good idea. Besides a better visual impact you can also better graduate the scale."
"The Barometer is an excellent tool for surveying."
"Very good."
"Good instrument."
"This new barometer is a good tool to visualize expectations."
"This is a very useful tool, thank you."

Negative Comment

"I do not see any improvement by using the barometer."

Comments on Problems with VAS

"No bar. Could not make this part of the survey work-so not completed."
"The cursor in 11.1 and 11.2 didn't function – error."
"There is difficult to move the slide bar."
"This feature does not work on my computer."
"Congratulations this time your system works."
"Business sentiment was not clickable."
"I was unable to move the barometer."

Table 1.4. Development of the Number of Clicks on the VAS in the Variables "Present Economic Situation" (Present) and "Economic Expectations" (Expect.) in Eight Consecutive Survey Waves (Balanced Panel of 60 Respondents)

Quarter	Median number of clicks		Mean number of clicks		Std. deviation of the mean		Maximum number of clicks	
	Present	*Expect.*	*Present*	*Expect.*	*Present*	*Expect.*	*Present*	*Expect.*
Q2/2005	3.0	6.0	4.1	5.4	3.9	3.2	18	10
Q3/2005	1.5	5.0	2.6	5.2	2.1	3.1	10	10
Q4/2005	1.0	6.0	2.3	5.9	2.1	3.0	13	10
Q1/2006	1.0	4.5	2.2	5.1	2.8	3.1	20	10
Q2/2006	1.0	4.0	2.1	4.6	2.1	3.0	13	10
Q3/2006	1.0	6.0	1.7	5.9	1.5	2.9	10	10
Q4/2006	1.0	5.0	1.5	5.6	1.2	2.9	8	10
Q1/2007	1.0	5.0	1.9	4.9	1.5	2.8	7	10

Table 1.5. Frequency of Inconsistent Responses between the Three-category Scale and the VAS

Quarter	Total Obs.	Present Economic Situation		Economic Expectations	
		%	Obs.	%	Obs.
Q2/2005	247	0.8%	2	2.8%	7
Q3/2005	289	1.7%	5	2.1%	6
Q4/2005	280	1.8%	5	3.2%	9
Q1/2006	293	0.7%	2	2.7%	8
Q2/2006	326	0.9%	3	0.9%	3
Q3/2006	315	2.2%	7	3.2%	10
Q4/2006	321	1.6%	5	4.0%	13
Q1/2007	322	1.2%	4	4.0%	13

Obs. = Number of observations

Table 1.6. Correlation Spearman's rho between the Three-category Scale and the VAS for the Variables "Present Economic Situation" (Present) and "Economic Expectations" (Expect.)

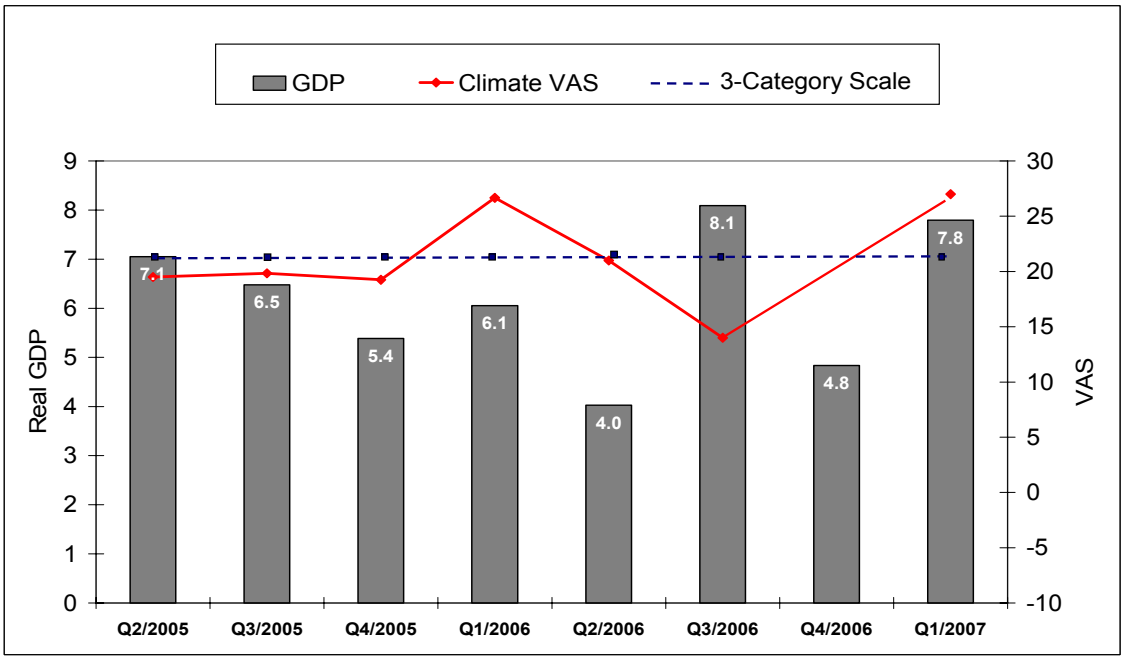
Quarter	Germany			USA			Japan		
	Obs.	Present	Expect.	*Obs.*	Present	Expect.	*Obs.*	Present	Expect.
Q2/2005	*17*	0.64	(0.40)	*14*	0.70	0.67	*11*	0.73	0.67
Q3/2005	*18*	0.66	0.74	*10*	(0.52)	0.91	*10*	0.65	0.64
Q4/2005	*19*	0.68	0.62	*15*	0.73	0.88	*14*	0.83	0.68
Q1/2006	*24*	0.82	0.59	*19*	0.78	0.58	*14*	0.67	0.61
Q2/2006	*25*	0.78	(0.33)	*13*	0.81	0.57	*14*	0.61	(0.47)
Q3/2006	*30*	0.53	0.73	*13*	0.75	(0.48)	*11*	0.71	(0.17)
Q4/2006	*30*	0.82	0.85	*11*	(0.60)	0.84	*13*	0.81	0.69
Q1/2007	*21*	0.49	0.71	*12*	0.88	0.80	*15*	(0.51)	0.69

Obs. = Number of observations
Not significant correlations in parenthesis

Figure 1.3. Time-trends of the Economic Climate Indicators* Derived with the Three-category Scale and the VAS for OECD Countries

* Economic Climate Indicator is calculated as the mean of the assessment of the present economic situation and economic expectations.

Figure 1.4. The Real GDP Growth in Ireland and the Business Climate Indicator Derived with the VAS and the Three-category Scale



Source: Real GDP data extracted on 2007/07/26 from OECD.Stat.

Figure 1.5. The Experimental Settings of the Test-retest Reliability Study

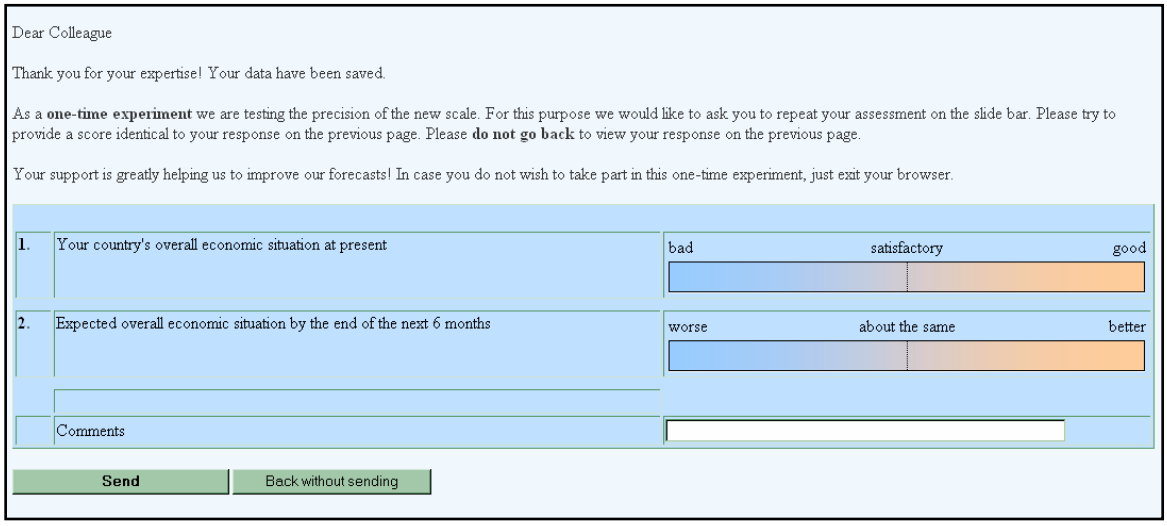


Table 1.7. Comparison of Participants and Non-participants in the Re-test Experiment Relating their Responses

Variables	Means		Test Statistics	
	Participants	Non-participants	t value	t-test significance Ho: Difference=0
Observations			**269**	**72**
Present economic situation	1.5	1.6	-0.42	0.68
Economic situation vs. last year	1.8	1.9	-1.75	0.08
Economic expectations	2.1	2.1	0.01	0.99
Capital expenditures	1.7	1.6	0.38	0.70
Capital expenditures vs. last year	1.8	2.0	-1.97	*0.05*
Capital expenditures expectations	2.1	2.0	0.68	0.50
Private consumption	1.6	1.7	-0.60	0.55
Private consumption vs. last year	1.8	2.0	-1.80	0.07
Private consumption expectations	2.1	2.1	0.03	0.98
Export expectations	1.8	1.5	2.74	*0.01*
Import expectations	1.6	1.5	1.14	0.25
Trade balance (surplus/deficit)	2.2	2.1	0.73	0.47
Inflation expectations	1.6	1.6	0.04	0.97
Inflation rate	4.7	4.3	0.36	0.72
Short-term interest rates	1.8	1.7	0.83	0.41
Value of the US$ expectations	2.1	2.1	-0.17	0.87
Stock market expectations	1.7	1.6	0.38	0.71
VAS Present economic situation	13.8	15.1	-0.45	0.65
VAS Economic Expectations	2.0	1.2	0.34	0.73

Each line represents the means in the respective variable and the results of a t-test on the equality of the means.

Table 1.8. Mean VAS Values in the Standard Questionnaire and the Re-test Experiment

VAS Variables	Response	Mean	Std. error of the mean	Std. Dev.	Clicks
Economic Situations	**Standard questionnaire**	13.8	1.2	20.3	1.9
	Re-test Experiment	13.8	1.3	20.4	1.4
Economic Expectations	**Standard questionnaire**	2.0	1.0	15.9	3.9
	Re-test Experiment	1.7	1.0	16.0	3.7

N=264 (5 respondents participated in the re-test experiment but have not provided an answer to the VAS question in the standard questionnaire. The means thus could be calculated only for 264 of 269 respondents.)

Table 1.9. Distribution of the Absolute Difference Between the Initial VAS Response in the Standard Questionnaire and the VAS Response in the Re-test Experiment

Absolute difference	Economic Situation				Economic Expectations			
	All respondents		Response away from an anchor*		All respondents		Response away from an anchor*	
	N	%	N	%	N	%	N	%
0	67	25.4	40	18.5	87	33.0	41	22.4
1	82	31.1	65	30.1	90	34.1	59	32.2
2	54	20.5	51	23.6	38	14.4	35	19.1
3	28	10.6	28	13.0	18	6.8	17	9.3
4	13	4.9	12	5.6	12	4.6	12	6.6
5	6	2.3	6	2.8	6	2.3	6	3.3
6	5	1.9	5	2.3	4	1.5	4	2.2
7	2	0.8	2	0.9	2	0.8	2	1.1
8	1	0.4	1	0.5	4	1.5	4	2.2
9	2	0.8	2	0.9	1	0.4	1	0.6
10	-	-	-	-	-	-	-	-
11	1	0.4	1	0.5	-	-	-	-
12	2	0.8	2	0.9	-	-	-	-
17	1	0.4	1	0.5	-	-	-	-
24	-	-	-	-	1	0.4	1	0.6
30	-	-	-	-	1	0.4	1	0.6
Total**	**264**	**100**	**216**	**100**	**264**	**100**	**183**	**100**

***** Respondents whose marker position in the initial VAS questions in the standard questionnaire was not on the middle anchor (+/-1) or on one of the extremes (+/-1). **5 respondents participated in the re-test experiment but have not provided an answer to the VAS question in the standard questionnaire. The RMSE thus could be calculated only for 264 of 269 respondents.

Table 1.10. The Mean Absolute Difference Between the VAS Response in the Standard Questionnaire and the Response in the Re-test Experiment for the Variables "Economic Situation" and "Economic Expectations"

Observations by marker position on the VAS	Obs.	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]		t-Test *one-tailed* *Ho: mean>0*
Present Economic Situation							
Total	264	1.81	0.13	2.18	1.55	2.07	t = 13.52
Close to an anchor [a)]	48	0.56	0.11	0.80	0.33	0.79	t = 4.89
Away from an anchor	216	2.09	0.16	2.29	1.78	2.39	t = 13.43
Economic Expectations							
Total	264	1.63	0.17	2.82	1.29	1.97	t = 9.39
Close to an anchor [a)]	81	0.49	0.07	0.63	0.35	0.63	t = 7.00
Away from an anchor	183	2.14	0.24	3.24	1.66	2.61	t = 8.91

a) The marker position on the slider was placed at one of the three anchors (the middle anchor or one of the two extremes) or close to them (+/-1 point of the scale).

Table 1.11. Factors Explaining Test-retest Reliability – Linear Regression Results of the Imprecision Interval on Explanatory Variables for the Variables "Economic Situation" and "Economic Expectations"

Factors relating to the imprecision interval	Economic Situation	Economic Expectations
Distance to an anchor (0-25)	0.021 **	0.032 ***
	(0.010)	*(0.013)*
Marker position at the middle anchor +/-2 (1-yes, 0-no)	-1.087 ***	-1.528 ***
	(0.392)	*(0.365)*
Frequency of participation (1-11 participations)	-0.079 **	-0.059
	(0.040)	*(0.051)*
Reminder (1-yes, 0-no)	-0.109	-0.140
	(0.292)	*(0.379)*
RMSE in the other VAS variable (0-30)	0.116 ***	0.182 **
	(0.046)	*(0.078)*
Difference to the marker position of the other VAS variable	0.005	0.011
	(0.009)	*(0.011)*
Constant	2.179	2.048
	(0.395)	*(0.533)*
R^2	*0.11*	*0.12*

Each column presents the results of a separate regression of the individual imprecision interval on the set of independent variables presented in the rows. Standard errors in parenthesis. N=264. **Asymptotic significance level< 0.05. ***Asymptotic significance level <0.01.

Table 1.12. Spearman's rho Between the VAS and Three-category Based Variables "Present Economic Situation" and "Economic Expectations" and Other Judgment and Expectations Variables of the Questionnaire

Variables from the questionnaire	Present Situation (3-Cat.)	Present Situation VAS	Economic Expect. (3-Cat.)	Economic Expect. VAS
Capital expenditures	0.02	0.06	0.68	0.53
	2316	*2145*	*2317*	*2143*
Private consumption	0.02	0.06	0.65	0.54
	2315	*2143*	*2316*	*2141*
Export expectations	0.11	0.12	0.31	0.27
	2322	*2150*	*2316*	*2148*
Import expectations	0.09	0.12	0.28	0.26
	2319	*2147*	*2313*	*2145*
Capital expenditures	0.69	0.63	-0.05	0.16
	2317	*2145*	*2310*	*2143*
Capital expenditures, vs. last	0.33	0.38	0.19	0.28
	2317	*2145*	*2311*	*2143*
Private consumption	0.62	0.59	-0.07	0.11
	2312	*2139*	*2304*	*2137*
Private consumption, vs. last	0.32	0.37	0.16	0.25
	2318	*2145*	*2312*	*2143*
Present growth of real GDP (%)	0.37	0.38	-0.01	0.14
	528	*483*	*527*	*483*
Problem: Government's	0.34	0.37	0.08	0.18
	1146	*1058*	*1144*	*1057*
Problem: Insufficient demand	0.40	0.40	-0.05	0.02
	1144	*1056*	*1142*	*1055*
Problem: Unemployment	0.22	0.23	0.00	0.05
	1145	*1057*	*1143*	*1056*
Problem: Inflation	0.05	0.05	0.09	0.07
	1145	*1057*	*1143*	*1056*
Problem: Lack of int.	0.22	0.25	0.07	0.15
	1143	*1055*	*1141*	*1054*
Problem: Trade barriers to	0.04	0.03	0.04	0.01
	1144	*1056*	*1142*	*1055*
Problem: Lack of skilled labour	-0.14	-0.13	0.07	0.01
	1143	*1055*	*1141*	*1054*
Problem: Public deficits	0.29	0.32	0.05	0.13
	1142	*1054*	*1140*	*1053*
Problem: Foreign debts	0.21	0.22	0.11	0.14
	1139	*1051*	*1137*	*1050*
Problem: Capital shortage	0.15	0.18	0.08	0.10
	1134	*1048*	*1132*	*1047*

a) The variable contains point forecasts. All other variables are measured by a 3-category scale.

Table 1.13. Cronbach's Alpha (Scale Reliability) of the Common Factor "Overall Economic Situation" and "Overall Economic Expectations"

Indicator / Items	Obs.	Sign	Item-test correlation	Average inter-item correlation	Cronbach's Alpha
Overall Economic Situation				0.33	**0.78**
Capital expenditures	2318	+	0.75	0.29	0.71
Capital expenditures, vs. to last year	2318	+	0.71	0.32	0.74
Private consumption	2312	+	0.72	0.31	0.73
Private consumption, vs. to last year	2319	+	0.69	0.33	0.75
Problem: Government's econ. policy	1146	+	0.53	0.36	0.77
Problem: Insufficient demand	1144	+	0.64	0.34	0.76
Real GDP growth (%)	528	+	0.54	0.34	0.76
Overall Economic Expectations				0.32	**0.68**
Capital expenditures expectations	2318	+	0.73	0.29	0.62
Private consumption expectations	2317	+	0.70	0.31	0.65
Export expectations	2324	+	0.68	0.33	0.66
Import expectations	2321	+	0.71	0.30	0.64

Obs. = Number of observations

Table 1.14. Measures of Association (Correlation Coefficient Spearman's rho and Coefficient of Determination R^2) between the Common Factor Variable and the Variable Present Economic Situation Measured by the Three-category Scale and the VAS

Quarter	Three-category Scale			VAS		
	Spearman's rho	R^2	Obs.	Spearman's rho	R^2	Obs.
Q2/2005	0.68	0.48	268	0.72	0.55	235
Q3/2005	0.67	0.48	264	0.72	0.53	239
Q4/2005	0.72	0.54	275	0.73	0.55	252
Q1/2006	0.60	0.41	314	0.60	0.42	296
Q2/2006	0.65	0.44	304	0.69	0.50	285
Q3/2006	0.62	0.43	300	0.57	0.41	286
Q4/2006	0.62	0.39	301	0.63	0.44	288
Q1/2007	0.65	0.44	303	0.62	0.42	275
Average	**0.65**	**0.45**		**0.66**	**0.48**	

Obs. = Number of observations

Table 1.15. Countries with More than Five Responses Quarterly (The Dataset for the ANOVA Analysis)

	2005			2006				2007	
Country	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Total**
Australia	7	0	0	0	0	0	0	0	**7**
Belgium	7	6	10	12	10	11	9	8	**73**
Brazil	6	0	0	5	0	6	0	8	**25**
Bulgaria	6	9	7	7	10	10	7	8	**64**
Chile	6	6	0	0	8	7	0	6	**33**
China	0	0	0	6	6	0	0	0	**12**
Czech Rep.	0	0	0	0	0	0	6	0	**6**
Finland	12	7	11	17	10	9	15	17	**98**
France	0	0	0	7	0	7	6	6	**26**
Germany	19	19	20	25	25	30	29	23	**190**
Greece	7	10	11	11	9	6	8	8	**70**
Italy	10	7	6	7	6	8	7	10	**61**
Japan	15	14	17	16	15	13	14	18	**122**
Mexico	0	0	6	0	0	7	0	0	**13**
Netherlands	8	8	9	10	11	9	8	6	**69**
Poland	6	8	0	6	9	9	7	7	**52**
Portugal	0	0	0	0	6	0	0	0	**6**
Russia	0	6	0	7	6	7	9	0	**35**
Slovakia	0	0	0	7	0	0	0	0	**7**
Slovenia	6	6	8	6	8	7	7	7	**55**
South Africa	8	7	6	7	8	9	7	7	**59**
Spain	9	7	6	9	10	9	8	7	**65**
Sweden	5	0	0	7	0	0	6	0	**18**
Switzerland	0	7	6	6	6	0	6	7	**38**
UK	7	0	0	0	0	0	0	0	**7**
USA	12	12	14	18	13	13	12	14	**108**
Total	**156**	**139**	**137**	**196**	**176**	**177**	**171**	**167**	**1,319**

Figure 1.6. Difference Between the Individual's Value and the Mean Country Value for a Particular Quarter as Function of the Country Mean Value

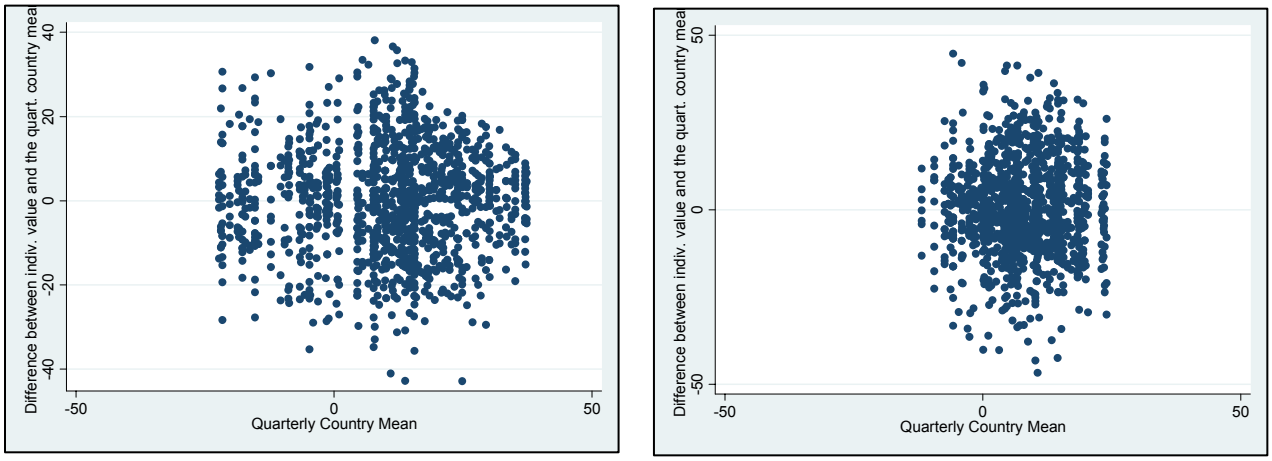


Table 1.16. One-way ANOVA results and the Intra-Class Correlation (ICC) for the Three-category Scale and the VAS

Quarter	Present Ec. Situation	SSB	DF (SSB)	SSW	DF (SSW)	F Ratio	ICC
Q2/2005	Categories	682.4	17	526.3	136	10.4	0.91
	VAS	33290.1	17	24355.0	118	9.5	0.90
Q3/2005	Categories	727.2	15	470.0	123	12.7	0.93
	VAS	34087.3	15	14925.3	109	16.6	0.94
Q4/2005	Categories	462.7	13	432.8	123	10.1	0.91
	VAS	23240.3	13	19982.4	109	9.8	0.91
Q1/2006	Categories	465.3	19	729.4	176	5.9	0.86
	VAS	28300.9	19	32450.9	164	7.5	0.88
Q2/2006	Categories	404.9	17	592.8	158	6.3	0.86
	VAS	21601.0	17	29165.4	147	6.4	0.86
Q3/2006	Categories	284.9	17	595.7	158	4.4	0.82
	VAS	13412.1	17	27521.3	151	4.3	0.81
Q4/2006	Categories	236.8	17	470.8	153	4.5	0.82
	VAS	16020.9	17	23295.4	147	5.9	0.86
Q1/2007	Categories	233.6	16	472.7	149	4.6	0.82
	VAS	23211.2	16	20069.0	136	9.8	0.91
Quarter	**Economic Expectations**	**SSB**	**DF (SSB)**	**SSW**	**DF (SSW)**	**F Ratio**	**ICC**
Q2/2005	Categories	194.0	17	695.7	136	2.2	0.69
	VAS	6741.1	17	22366.7	118	2.1	0.68
Q3/2005	Categories	201.0	15	745.5	123	2.2	0.69
	VAS	4702.3	15	15974.4	110	2.2	0.68
Q4/2005	Categories	344.2	13	624.4	123	5.2	0.84
	VAS	6080.2	13	16423.0	110	3.1	0.76
Q1/2006	Categories	270.4	19	808.7	173	3.0	0.75
	VAS	8490.2	19	28091.1	164	2.6	0.72
Q2/2006	Categories	280.0	17	759.9	158	3.4	0.77
	VAS	8610.6	17	27663.3	147	2.7	0.73
Q3/2006	Categories	373.1	17	774.9	157	4.4	0.82
	VAS	9398.0	17	31615.0	152	2.7	0.73
Q4/2006	Categories	344.6	17	850.8	153	3.6	0.78
	VAS	11904.3	17	24835.8	147	4.1	0.81
Q1/2007	Categories	239.2	16	738.9	149	3.0	0.75
	VAS	10610.8	16	24974.2	136	3.6	0.78

SSB – Sum of squares between groups
DF (SSB) – Degree of freedom between groups
SSW – Sum of squares within groups
DF (SSW) – Degree of freedom within groups
ICC – Intra-class correlation for country mean $ICC = \frac{(SSB/DF(SSB))}{SSB/DF(SSB) + (SSW/DF(SSW))}$

Table 1.17. Intra-Class Correlations (ICC) of the Three-category Scale and the VAS for Each Quarter and on Average

Quarter	Present Economic Situation		Economic Expectations	
	Three-category Scale	VAS	Three-category Scale	VAS
Q2/2005	0.91	0.90	0.69	0.68
Q3/2005	0.93	0.94	0.69	0.68
Q4/2005	0.91	0.91	0.84	0.76
Q1/2006	0.86	0.88	0.75	0.72
Q2/2006	0.86	0.86	0.77	0.73
Q3/2006	0.82	0.81	0.82	0.73
Q4/2006	0.82	0.86	0.78	0.81
Q1/2007	0.82	0.91	0.75	0.78
Average ICC	**0.87**	**0.88**	**0.76**	**0.74**

Chapter 2

Validity of Economic Indicators Derived with the Visual Analog Scale

Abstract

This chapter evaluates the validity of the indicators derived with the visual analog scale. Responses elicited with the VAS are qualitative data and are at the same time close to an interval-scale measurement allowing not only the construction of direct indicators of business expectations and business present situation, but also the estimation various dispersion measures and uncertainty proxies at the cross-sectional level. The time-series of business confidence, dispersion measures and uncertainty proxies are evaluated for their empirical relationship to economic performance in the German manufacturing sector in the time-period September/2005 to July/2008. The data set is based on the business tendency survey in the German manufacturing sector and contains 46,022 observations from 34 consecutive monthly surveys.

1. Introduction and Background

Business tendency surveys (BTS) collect up-to-date information on present and future economic activity. In contrast to the official statistics, which rely on quantitative information, BTS mostly use qualitative assessments, as they are fast and easy to collect. These qualitative assessments are usually used to track and to forecast the macroeconomic performance in a sector or the overall economy. In particular, business expectations have proved to be a valid indicator of future economic growth. The present paper evaluates the validity of various indicators derived with the new qualitative measurement instrument of economic expectations – the visual analog scale (VAS).

Validity identifies the degree to which a test measures what it is supposed to measure, or the degree to which theory and empirical evidence support the interpretations of the test scores (APA, 2004). Establishing validity is a complex task. Historically, validity has been divided into different validity types (APA, 1966): content validity, construct validity, criterion-related validity. Later methodologists argued for a unified framework seeing the other types of validity as subtypes of construct validity. This view predominates today, although researchers still validate the inferences in multiple ways (Messik, 1989; Adcock and Collier, 2001; Westen and Rosenthal, 2005): relevant content, based on sound theory or rationale, internally consistent items, criterion-related evidence and validity generalization across populations and time. The theory and rationale behind the qualitative measurement of economic expectations was established when business surveys were in their early stages by Anderson (1952) and Theil (1955). Reliability and internal consistency of the VAS is discussed in the previous chapter. Consequently, the present paper will not address all validity issues but will focus on the criterion-related evidence in particular.

Criterion-related validity reflects to which extent a measurement correlates with known and accepted criteria that measure the same content. Ideally these criteria are direct, objective outcomes of what is being measured. So, business expectations

should be compared with the company's realizations in form of production output or profits. However, these direct objective measures are usually unavailable, so that other criteria have to be chosen that are closely associated to them.

Criterion evidence can be established in two ways, using the concurrent and the predictive design. In this study both approaches are applied. The economic literature has mainly focused on business confidence indexes to track macroeconomic outcomes, particularly GDP (Nerlove, 1983, Bandholz et al., 2003, Claveria, Pons and Ramos, 2007) or production index growth rates (Breitung and Jagodzinski, 2001, Fritsche and Stephan, 2002, Hüfner and Schröder, 2002, Benner and Meier, 2004, Dreger and Schumacher, 2005, Wohlrabe and Robinzonov, 2008). In line with these studies, the concurrent validity is established by evaluating whether the proposed VAS measures exhibit the same direction and magnitude of correlation with the criteria as does the established measure of business expectations, which is based on the three-category scale.[1] As a direct measure of the company's realizations is not available, the used criterion is the monthly year-on-year growth rate of the production index in the German manufacturing sector. In research on consumer confidence, studies also evaluated the ability of individual responses to predict subsequent individual outcomes in re-interviews (Dominitz, 2001, Dominitz and Manski, 2004). This question was the subject of analysis also in earlier studies on business surveys (Theil, 1955, Nerlove, 1983, Gourieroux and Pradel, 1986, Ivaldi, 1992), although the authors generally addressed this question under the heading "rationality of expectations". Predictive validity in the present study is also established by comparing the business expectations with the assessment of the business situation in subsequent surveys.

Not only the direct measures of business expectations are evaluated in the study. VAS data allow the estimation of higher moments of distribution at the cross-sectional level, such as the dispersion of business expectations and kurtosis. Although VAS collects rather qualitative information, its scale properties appear to be somewhere between an ordinal and an interval measurement (Reips and Funke,

[1] The measurement procedure of business surveys on economic expectations has been harmonized by the European Commission in the European Union to three-category rating questions.

2008, Butler, 1997), because VAS responses can be converted to a continuum of numerical values (as a rule on a 1-to-100 scale).

In surveys of professional forecasters the dispersion of economic expectations is used as proxy for uncertainty. The assumption required for the validity of using the dispersion as proxy for uncertainty is that the interpersonal dispersion measure is a good approximation to the intrapersonal forecast dispersion or uncertainty about the future development of the economy or the company (Zarnowith and Labros, 1987). Although some authors doubt the validity of the dispersion measures as proxy for uncertainty (Rich and Tracy, 2006, Mitchel et al., 2005, Lahiri and Liu, 2006, Doepke and Fritsche, 2006, Boero et al., 2007), there are numerous studies supporting the use of the forecast dispersion as proxy for macro-economic uncertainty (Bomberger, 1996, Batchelor and Dua, 1996, Giordani and Soderlind, 2003, Linden, 2003). However, there are numerous arguments that dispersion of business expectations[2] is important in its own right and not just as proxy for uncertainty (Mankiw, Reis and Wolfers, 2003, Doepke and Fritsche, 2006, Giordani and Soderlind, 2003).

Business expectations of companies depend on both, the macro-economic environment and the individual company's characteristics. Consequently, the dispersion of business expectations includes two components: uncertainty about the path of the macro-economic development and heterogeneity due to company's characteristics. Empirical evidence suggests that the heterogeneity of business expectations has cyclical patterns. Business expectations are less heterogeneous in times of a strong economy, i.e. also economically weak businesses perform well. In contrast, in times of weak economic growth, business expectations diverge resulting

[2] Doepke and Fritsche (2006) argue that examining the heterogeneity of economic expectations is an important question as it is commonly assumed in economic analyses and models that economic agents share identical beliefs about the structure of the economy, which may not be true. The authors summarize the reasons for heterogeneity relating to future economic development among economic agents: (a) Forecasters may have different knowledge about the current state of the economy at the time the forecast is made. (b) Forecasters may rely on different assumptions on the path of variables, which are in most cases exogenous to the forecasting process, e.g., oil prices or the course of economic policy. (c) Forecasters may have different ideas on the appropriate model of the economy and its adjustment dynamics. (d) And some strategic forecasting behavior can also not be ruled out, if the forecasters are not anonymous.

in higher heterogeneity, i.e. strong businesses continue to perform well, but business expectations of economically weak companies worsen.

Along with dispersion measures, VAS offers further proxies of uncertainty. The proportion of neutral responses (responses at the middle anchor of the VAS), which may indicate "epistemic uncertainty", according to the Bruine de Bruin et al. (2000) definition. Bruine de Bruin et al. argue that a 50% response or heaping at the middle reflects "epistemic" uncertainty, and only seldom represents the true view of the respondent. The anchor rather functions as a surrogate for a "don't know" or "unsure about the future development" option, even if a "don't know" option is available within the question. However, not all neutral responses reflect epistemic uncertainty. A particular proportion of neutral responses may reflect the true views of the respondents. This proportion is not known and the data are used under the assumption that the proportion of responses reflecting the true view remains stable over time relative to the proportion of responses that reflect epistemic uncertainty.

The role of uncertainty in influencing economic behavior of agents, such as investment decisions or monetary-policy, has been recognized early (Knight, 1921) and is nowadays widely accepted. Risk-aversive economic agents may delay investment and consumption decisions if they are uncertain about the future economic development. Consequently, high macroeconomic uncertainty may negatively influence an optimal allocation of resources and lead to welfare losses. The most common empirical finding is that economic uncertainty and heterogeneity of economic expectations is negatively related to the economic cycle, i.e. higher uncertainty is associated with lower output growth (Ramey and Ramey, 1995, Sepulveda-Umanzor, 2003, Mitchell et al., 2005, Doepke and Fritsche, 2006), and lower investments (Guiso and Parigi, 1999). For the German economy, Doepke and Fritsche (2006) based on a panel of German professional forecasts for 1970–2004 analyzed the dispersion of growth and inflation forecasts. They found that forecast dispersion varies over time and is particularly high before and during recessions and in the early upswing. Consequently, it is expected that the dispersion of business expectations and the "epistemic" uncertainty will be also generally negatively

correlated with the reference series of economic performance in the manufacturing sector.

The plan of the chapter is as follows. The study compares the business expectations derived with the VAS and with the three-category scale, and evaluates their concurrent and predictive validity with respect to the chosen criteria. Also additional cyclical indicators are derived from the VAS responses and evaluated for their empirical relationship to the economic growth in the German manufacturing sector. The data set for the time period September/2005 to June/2008 and the benchmark criteria are described in the following section. This section also discusses how the measures of dispersion and heterogeneity of business expectations are constructed and whether there are statistically significant differences among these measures. The empirical results are described in section 3. The paper closes with a discussion.

2. Data

2.1. The Business Survey in the German Manufacturing Sector

The VAS was implemented within a monthly Business Tendency Survey (BTS) in the German manufacturing sector. The survey elicits information from companies on a number of business variables, such as the assessment of the present business situation and business expectations, evaluation of prices, sales, exports and other relevant business parameters (Figure 2.1 contains a sample paper questionnaire). The panel involves business officials with specializations within their companies in management, finance, and other strategic business functions. The survey participation is absolutely voluntary and derives entirely from the interest in the survey results, as no other compensation is offered. Becker and Wohlrabe (2008) provide a detailed description of the Ifo Business Survey micro data.

The overall Ifo BTS sample, including paper survey participants contained by 12/2006 2,622 sample units – approximately 1 percent of German enterprises in the manufacturing sector. The companies covered by the sample include small, medium-sized and large companies. As BTSs are designed to forecast the business cycle in a sector with respect to output growth, particularly big economic players are involved in the survey (see Annex, Table 2.1). In the present study, however, only the Internet survey responses are used. The Internet sample contains approximately 0.5 percent of all companies in the German manufacturing sector.

The dataset of individual responses covers 34 survey waves, beginning with the September 2005 survey and continuing to the June 2008 survey. Each respondent has an identification number, so that their individual responses can be tracked over time. The total number of responses in the dataset is 46,022. The data of the study comprise a conventional incomplete panel dataset, as the sample changes from time to time and not every respondent provides an expertise every month. Only 157 respondents are ever-present in all 34 survey rounds. However, the mean number of survey participations during the 34-months period is 27, indicating that the panel is sufficiently stable. The majority of the following analyses were thus conducted using the whole data-set.

The variables under analysis are the 6-month expectations over the general business conditions and the assessment of the present business situation. In the present experiment responses on both variables were elicited twice – with the three-category scale and with the VAS. A screen-shot of the VAS is displayed in the Annex (Figure 2.2). Figure 2.3 shows the distribution of business expectations on the three-category scale and on the VAS. It illustrates the typical peak around the central anchor of the VAS scale. This proportion of responses is in the following analysis used as proxy for “epistemic” uncertainty.

2.2. Production Index for the German Manufacturing Sector

The performance of the economy is proxied by the monthly year-on-year growth rate of the production index (year 2000=100) for the German manufacturing sector. Business survey respondents are asked to respond to the questions on business expectations without taking into account differences in the length of months or seasonal fluctuations. However, their mental adjustment for working day variations and seasonal fluctuations may be different from the computational methods of the German Federal Statistical Office. For that reason two measures of the production index growth (PI) are used in the paper: Value adjusted for working day variations, not seasonally adjusted (nsa) and value adjusted for working day and seasonal variations (sa). Growth rates of the production index are calculated as percentage change of the index compared to the same month of the previous year. The German Statistical Office continuously revises the production indices backward. The data used in the study were retrieved in August 2008 and the time-series are summarized in Table 2.2.

2.3. Survey-based Business Climate Indicators (BCI)

2.3.1. Business Climate Indicator (BCI) Based on Three-category Responses

The second type of proxy for the economic performance in the German manufacturing sector is the survey-based business climate indicator (BCI). It is constructed from the two variables "assessment of the present business situation" and "business expectations for the next six months". The question wording usually is: "*This company's overall business situation at present?"* and *"Expected overall business situation by the end of the next 6 months?"* The idea behind such business climate questions – also called business sentiments or business confidence – is that polled experts assess the overall business situation in their company by taking into

account all the aspects they regard as important. Although no precise information on levels of output or profits is usually asked, BCI can be used to predict changes in production and, for that reason, it is particularly useful for the forecast and analysis of the business cycle (see Annex, Figure 2.4). The three-category responses are weighted with the company's economic weight, measured by the number of employees.[3] After weighting, balance statistics (fraction of positive responses minus the fraction of negative responses) for various product groups are calculated. The product group breakdown corresponds to the German Statistical Office branch classification ('WZ2003'). The product group results and later the branches are weighted by their shares in gross value added, according to the official statistics.

2.3.2. Business Climate Indicator Based on the VAS Responses (BCI-VAS)

A BCI can be also constructed from the VAS questions on the present business situation and business expectations and be used as a reference measure (BCI-VAS). The BCI-VAS is constructed as arithmetic mean of the assessment of the present business situation and business expectations. The weighting procedure corresponds to the methodology of the three-category scale responses described above: the company's responses are weighted by the number of employees, product groups and branches are weighted by their shares in gross value added.

The BCI and the BCI-VAS have the favorable characteristic that they are constructed from the same responses as the dispersion measures and the uncertainty proxy, while the two production indices are based on quantitative information collected by the German Statistical Office and include also data from companies not covered by the survey. In the long-term (1996 – 2008) the correlation of the seasonally adjusted BCI with the Production Index (sa) is 0.84 (see Annex, Figure 2.4). The BCI and the BCI-VAS are also highly correlated (see Annex, Figure 2.5).

[3] The number of employees is not used directly as a weighting factor, but by using a logarithmic transformation. In this way, responses from a large participant are prevented from dominating the aggregated results of its product group; at the same time, this weighting scheme acts to correct the under-representation of small business units in the survey.

2.4. Dispersion of Business Expectations and the "Epistemic" Uncertainty

The uncertainty measures suggested in the literature can be divided into direct survey-based measures and indirect survey-based measures.[4] The direct survey-based measures of uncertainty are derived by eliciting a complete probability distribution of a point forecast, as it is done in surveys of professional forecasters (SPF). [5] Indirect survey-based measures of uncertainty use the cross-sectional dispersion of individuals' expectations or point forecasts from surveys. Both measures of forecast uncertainty have specific advantages and limitations. The direct measure of uncertainty in form of a complete probability distribution of a point forecast elicited in SPFs is theoretically appealing and has several desirable features (Manski, 2004). The standard deviation of an individual histogram is associated with the uncertainty of the individual forecaster. The variance in the aggregate histogram incorporates both individual uncertainty and heterogeneity of expectations. The complete probability distribution provides a well-defined numerical scale of responses and makes it possible to compare responses across individuals. However, probabilistic expectations have been also critisized[6] (Boero et al., 2007, Engelberg et al., 2006, Clements, 2007). Furthermore, questions eliciting probabilistic expectations are rare, applicable to economists or people familiar with probability distributions, time-demanding and difficult to apply in business and household surveys as well as in panel settings (Boero et al., 2007, Mitchell et al., 2005). Consequently, popular

[4] Time-series based measures proxy uncertainty ex-post by the root mean square forecast error made by a rational forecaster. These measures of uncertainty are not discussed in the present study.

[5] In Surveys of Professional Forecasters (SPF) respondents are macroeconomic forecasters who are polled about future GDP growth and inflation and provide both point forecasts and the probabilities that the point forecast will lie in particular intervals. The best known SPF surveys are the US Federal Reserve Bank in Philadelphia SPF, the European Central Bank SPF and the Bank of England Survey of External Forecasters.

[6] Probabilistic measures tend to cause a high cognitive effort to respondents, particularly in panel settings and in surveys of non-economists. Some respondents may have no incentive to answer the questions carefully, nor the ability to deal with numbers (Boero et al., 2007). Furthermore, heaping of rounded and 50% values poses a serious empirical problem in surveys of non-economists (Guiso, Tiseno and Winter, 2005). In case, bins are provided to respondents, they may feel obliged to fill them with probabilities, even if they do not have an already defined probability distribution in mind. It has been also hypothesized by researchers that SPF forecasters might give more careful consideration to their point predictions than to their probabilistic forecasts (Engelberg et al., 2007). Furthermore, in a variety of surveys researchers observed a general tendency of respondents to be optimistic, i.e., to report high probabilities for positive and low probabilities for negative events (Guiso, Tiseno and Winter, 2005, Boero et al., 2007, Engelberg et al., 2006, Clements, 2007).

approaches to proxy uncertainty include cross-sectional measures, based on the dispersion of individuals' expectations or point forecasts in surveys. The dispersion of business expectations is considered to be a useful uncertainty proxy, given the assumption laid by Zarnowith and Labros (1987) holds, that the interpersonal dispersion is a good approximation to the intrapersonal forecast dispersion. Dispersion of economic expectations has the advantage of being readily available in almost all kind of surveys eliciting responses on interval or ordinal scales. In BTS, however, meaningful dispersion measures have not been available thus far, as business expectations are elicited with three-category scales within the European Union. VAS data, in contrast, allow the estimation of higher moments of distribution at the cross-sectional level, such as the dispersion of business expectations, kurtosis and skewness, because the scale properties of the VAS appear to be between an ordinal and an interval scale measurement.

Several measures of dispersion are tested in the study for the presence of a cyclical pattern: The first measure is the standard deviation of VAS business expectations (excluding the responses on the central anchor of the scale). The standard deviation includes both, a proxy of intrapersonal uncertainty and heterogeneity of business expectations. The cyclical behavior of both components is expected to be similar: intrapersonal uncertainty and heterogeneity of business expectations are higher in the early upswing and when economic growth falls, and decrease during high and stable economic growth periods.

The second measure of dispersion is kurtosis of the VAS business expectations (also calculated excluding the responses on the central anchor of the scale). A normal random variable has a kurtosis of 3 irrespective of its standard deviation. Comparing two PDFs relating their kurtosis, it is impossible to conclude that the PDF with the lower kurtosis has higher standard deviation or vice versa. Doepke and Fritsche (2006) examined in their study the distribution of point forecasts for economic growth and inflation. According to their hypothesis, a significant kurtosis above 3 indicates that the forecasters are very close to each other. Kurtosis is consequently also a measure of heterogeneity and uncertainty, although it may contain a different information set than the standard deviation measure.

There are further dispersion measures suggested in the literature: inter-quartile range (Giordani and Soderlind, 2003, Doepke and Fritsche, 2006) and skewness (Doepke and Fritsche, 2006). Giordani and Soderlind (2003) have chosen a dispersion measure that is robust to outliers and used the quasi-standard deviation, equal to half the difference between the 16th and 84th percentile of the sample of point forecasts. With normally distributed data this calculation delivers the standard deviation and prevents the measure from being overly sensitive to extreme observations. For the same reason Doepke and Fritsche (2006) use in addition to the standard deviation of the forecasts the inter-quartile range. Furthermore, they examine each year's forecast to determine whether there is a consensus among forecasters, by testing for the skewness. According to their hypothesis, if the distribution is significantly skewed, consensus among forecasters is rejected.

In common with other studies on uncertainty in economic expectations, measures of dispersion are compared to alternative uncertainty proxies. An alternative measure of uncertainty is the proportion of neutral responses (responses on the middle anchor of the VAS) or "epistemic" uncertainty according to Bruine de Bruin et al. definition (2000). If the dispersion measures of business expectations are useful proxies of uncertainty, one would expect a significant correlation between the dispersion measures and the "epistemic" uncertainty. Accordingly, a positive correlation is expected between the standard deviation, quasi-standard deviation and the inter-quartile range of VAS business expectations and the "epistemic" uncertainty and a negative relationship between the kurtosis of VAS business expectations and "epistemic" uncertainty.

The dispersion measures and "epistemic" uncertainty are computed for the German manufacturing sector at the cross-sectional level in the time period September/2005 to June/2008. The resulting time-series are summarized in the Annex (Table 2.2), so that all following calculations can be easily replicated.

The dispersion measures of business expectations and the "epistemic" uncertainty measures appear to be closely related to each other (see Annex, Table 2.4). In the respective time-period the "epistemic" uncertainty correlates with the dispersion of

business expectations by 0.66 ($p<0.05$) and with the kurtosis by -0.54 ($p<0.05$). The relationship between the two dispersion measures of VAS business expectations and the "epistemic" uncertainty is illustrated in Figure 2.6. Using robust measures of dispersion such as the inter-quartile range and the quasi-standard deviation decreases the correlation with the epistemic uncertainty.

As discussed earlier, the dispersion measures of business expectations include not only information on macro-economic uncertainty, but also reflect a natural heterogeneity of business expectations caused by different characteristics of individual companies. The heterogeneity of business expectations may also have cyclical properties. To extract the heterogeneity component, each dispersion measure of business expectations was regressed on the "epistemic" uncertainty. Then the residuals of each regression were calculated. The residual contains the information share of the dispersion measures that is not related to the "epistemic" uncertainty and can be used as proxy for the heterogeneity of business expectations. There were three dispersion measures that were found to be correlated with the "epistemic" uncertainty: Dispersion of business expectation, kurtosis and the inter-quartile range. Consequently three regressions were performed and the residuals are used in the following as heterogeneity measures (see Table 2.4).

In contrast to the dispersion measures of business expectations, none of the dispersion measures of the variable "present business situation" is significantly correlated to the proportion of neutral responses in the assessment of the present business situation. This result is intuitive, as the present business situation of the company is usually known. The uncertainty about the present state is consequently low or even absent. The dispersion of responses reflects the pure firm-specific heterogeneity and the proportion of neutral responses has a higher share of "true" values than the proportion of neutral responses in business expectations. The dispersion of business expectations, in contrast, contains both, firm-specific heterogeneity and inter-personal disagreement about future macro-economic conditions. These results provide additional evidence that the dispersion of business expectations can be used as a proxy for uncertainty. A side product of the above exercise is the proof that there is "epistemic" uncertainty in neutral business

expectations, while in the assessments of the present business situation the proportion of responses reflecting "epistemic" uncertainty is lower.

3. Results

3.1. Predictive Validity: Business Expectations Measured by the VAS and the Three-category Scale and the Assessment of the Present Business Situation in Subsequent Surveys

The predictive validity is established by analyzing the relationship between business expectations and the assessment of the business situation in the subsequent surveys. This relationship is explored in several ways: at the micro-level – regression of the business situation on lagged business expectation and a correlation analysis, and at the aggregated level by a regression of the time-series of business situation on the time-series of business expectations, considering different lags.

At a first step, the assessments of the business situation were regressed on lagged business expectations. Several lags of business expectations were taken into consideration, to account for the possibility that the realizations of the six month business expectations occur earlier and last longer than six months. The following variables were added as controls to the regression equation: company size class measured by the number of employees and time-dummies to account for shocks or aggregate effects that change over time (for example the US sub-prime credit crisis). The regressions were performed by panel fixed effects (with indicator variables for each observation and allowing for robust standard errors) to control for further unobserved variables, as for example the characteristics of the survey respondent, such as ability or firm specifics that do not change over time. Whether a random effect model that provides more efficient estimators can be used was evaluated with the Hausman test (1979) that specifies whether the random effects estimate is insignificantly different from the fixed effects estimate. However, the Hausman test

confirmed that there is an omitted variable that is correlated with one of the independent variables and the random effect estimator would be biased. For the VAS, a linear OLS regression with fixed effects can be easily applied. For the data elicited with the three-category scale, an ordered probit regression is recommendable. However, a fixed effects ordered probit approach that estimates the coefficient of the lagged expectations and the individual time-invariant parameters (like in the linear fixed effect model) is infeasible, as the model parameters suffer from the "incidental parameters problem" (Wooldridge 2002: 274) and the estimation of the incidental parameters along with the coefficients of the independent variables produces inconsistent results. As there is a continuous latent variable underlying the categorical assessment, one option would be to run a linear fixed effects model also for the ordinal data. Another option would be to run and ordered probit, ignoring the presence of the omitted individual effect. Both options appear unsatisfactory. Furthermore, within a three-category scale more than 50 percent of all assessments of the present business situation and more than 75 percent of responses about the business expectations fall in the neutral category. There is also a low changeover across categories over subsequent survey waves. On average, the most frequent transition state over two subsequent surveys is when respondents remain in the neutral category (65 percent of respondents opt for the neutral category over two subsequent surveys). For the assessment of the present business situation this share amounts 45 percent. Consequently, the relationship between the lagged three-category based business expectations and the three-category based assessments of the business situation may be overestimated due to the lack of variability in responses and the dominating proportion of neutral responses. Consequently, the analysis was undertaken only for the VAS data. The regressions were performed at the balanced panel under the assumption that the missing observations are random.[7] As 164 participants responded regularly during the 34 months period the data base contained all over 5,576 observations. The results are displayed in the Annex (Table 2.5). The results demonstrate that business expectations measured by VAS are highly

[7] Observations may be missing for many reasons (bankruptcy, new companies are sampled, contact persons quit companies, or are unwilling to participate in one or several survey rounds due to various reasons). The most common reason is that respondents are not able to participate due to vacations or time-constraints. These reasons are assumed to be uncorrelated with the level of their business confidence. However, the results do not substantially differ from the fixed effects regression on an unbalanced panel.

significant predictors of the future business situation and forecast the future business situation up to nine months ahead.

To compare the predictive validity of the VAS based business expectations and the three-category based business expectations, a correlation coefficient for ordinal data, the Spearman's rho is calculated to measure the relation of business expectations and the future assessment of the business situation for both scales and different lags. The results of the correlation analysis are displayed in the Annex (Table 2.6). The Spearman's rho is considerably higher for the VAS, ranging between 0.53 and 0.60, than for the three-category scale, where it ranges between 0.21 and 0.25. These results are indicating that the VAS displays a higher predictive validity in the micro-data. The correlation coefficients are tested on equality using the approximation test suggested by Cohen and Cohen (1983). The rank correlation of the lagged expectations and the business situation is significantly higher when the variables are measured by the VAS than by the three-category scale.

Another perspective on predictive validity can be obtained by looking at the aggregate time-series of business expectations and business situation. The indicators of business situation and the six months business expectations are derived by both, applying the weighting procedure described in section 2.3 and without weighting of the individual responses. The relation between the aggregated business situation and lagged aggregated business expectations is explored by the rank correlation coefficient Spearman's rho analysis and the R^2 from a regression analysis. The results of the Spearman's rho correlation between the time-series of aggregated business situation and aggregated business expectations are summarized for different lags in the Annex (Table 2.7). The correlation between the time-series of business expectations and business situation derived with VAS, both weighted and unweighted, are considerably higher than the correlation of the same time-series derived with the three-category scale. The correlation of the VAS business expectations with the future business situation is significantly higher than zero for up to six months ahead. The correlation coefficients of the indicators of business situation and lagged business expectations derived with the three-category scale appear not be significantly different from zero, except for the second lag. The

correlation measures of the unweighted time-series are higher than of the weighted time-series. This may be explained by the fact that the larger the company, the more difficult it is to forecast its future business state. During the aggregation procedure larger companies receive higher weights. The resulted weighted time-series of business expectations and future business situation are consequently less strongly correlated compared to the unweighted data.

Alternatively the indicator of business situation was regressed on the indicator of the six months business expectations considering different lags. A separate regression was performed for each of the nine lags. The determination coefficient R^2 of the regression is displayed in the Annex (Table 2.8). The results are very similar to the correlation analysis. Lagged time-series of business expectations explain a higher proportion of the variation of the business situation time-series if the time-series are derived with the VAS. Figure 2.7 illustrates more vividly the Spearman's rho and the R^2 for different lags of aggregated business expectations. The aggregated VAS business expectations are a better predictor of the VAS business situation in successive months, although the predictive power diminishes with each month. The predictive power of business expectations indicator derived with the three-category scale is much lower, although the R^2 appears to increase somewhat over time, it does not become considerably higher than the measures of the VAS.

In summary, the business expectations derived with the VAS display a higher predictive validity in the micro-data and in the aggregated time-series. The next chapter will establish validity of the VAS indicators applying the concurrent design.

3.2. Concurrent Validity of the Indicators Derived with the VAS

3.2.1. Business Climate Indicators (BCI and BCI-VAS) and the Economic Growth in the German Manufacturing Sector

In the time period September/2005 to June/2008 the correlation of the BCI with the PI(nsa) is, at 0.44, lower than it is in the long-term run but still significant. The BCI-VAS correlates with the PI(nsa) similarly strong, by 0.37. The correlation of the BCI and BCI-VAS is somewhat higher with the seasonally adjusted production index PI(sa) than with the not seasonally adjusted PI(nsa). The correlation between the BCI and BCI-VAS in the time-span under analysis is, at 0.96, very high and significant. These results are summarized in the Annex (Table 2.3).

The time-span of 34 months is too short for drawing any conclusions about the performance of the business confidence indicators in forecasting the business cycle in the German manufacturing sector or about their leading properties. However, the overall evidence suggests that all four measures of economic performance in the German manufacturing sector are significantly correlated, are reflecting the same business cycle pattern. In the following all four measures (PI(nsa), Pi(sa), BCI and BCI-VAS) are used as reference series for economic performance in the German manufacturing sector.

3.2.2. Dispersion of Business Expectations and "Epistemic Uncertainty" and the Economic Growth in the German Manufacturing Sector

There is an empirical relationship between the rate of change of the industrial production index and the standard deviation of business expectations. Figure 2.8 in the Annex shows the changes of the standard deviation of VAS business expectations and the PI(sa) over the experimental period. The standard deviation of business expectations is higher in periods of low economic growth, and particularly high during the economic expansion in 2005. It appears to be generally lower during

periods of stable growth, such as in the time-span December/2006 to February/2007. The standard deviation of business expectations increases as PI growth-rate falls, as for example in September 2006, in April 2007, in July 2007, the begin of the US sub-prime credit crisis and in May 2008.

The other dispersion measures are also negatively related with the indicators of economic growth except kurtosis, as expected, is positively correlated with the PI and the BCI and BCI-VAS (see Annex, Table 2.9). However, only "epistemic" uncertainty, standard deviation of business expectations and kurtosis are significantly correlated with the PI(sa/nsa), by -0.44, -0.50 and 0.47 respectively. The heterogeneity of business expectations is not significantly correlated with the PI(sa/nsa), but with the BCI and the BCI-VAS, by -0.51 and –0.45 respectively. Nevertheless, this result indicates that business expectations are less heterogeneous in times of high economic growth. In the early upswing and as economic growth falls the heterogeneity of business expectations increases. The cyclical characteristics of the heterogeneity measure is similar to the cyclical characteristics of the "epistemic" uncertainty, although the two indicators are entirely independent (see Table 2.4 and Table 2.9).

A further interesting finding is that skewness of the VAS business expectations appears to be procyclical: significantly negative during low-growth periods. Figure 2.9 visualizes the changes in the distribution of business expectations over the observed time-period. The distribution shifts from the negative side of the VAS and a negative skewness to the positive side of the VAS and a skewness not significantly different from zero. Although, skewness appears to be significantly positive correlated only with the BCI-VAS, the Figure 2.9 demonstrates that there is some cyclical behavior. It would be interesting to observe the skewness of the distribution in the cyclical trough and the early up-swing.

Not only the VAS business expectations contain valuable cyclical information and can help to forecast economic growth, further indicators may be derived from the VAS responses on the present business situation. Table 2.10 in the Annex

summarizes the various additional indicators and shows their empirical relationship to the four indicators of economic growth in the German manufacturing sector.

The proportion of neutral responses in the VAS business situation is negatively related to the economic growth, and significantly correlated with the PI(sa) and PI(nsa). But, the correlation are lower than the correlations of the "epistemic" uncertainty measure derived from business expectations. This result indicates, as discussed earlier, that in the neutral assessments of the present business situation the proportion of responses reflecting uncertainty is lower than in responses on business expectations.

The standard deviation of the VAS business situation is significantly negatively correlated, and adequately, the kurtosis is positively correlated with the economic performance. As in the case of business expectations, these results indicate that the heterogeneity across companies increases when economic growth rates are low.

The skewness of the VAS business situation is significantly negatively correlated with all four economic performance measures by, meaning that the distribution of the VAS responses on the present business situation is negatively skewed when economic growth is high. This pattern is contrary to the cyclical properties of the skewness of the VAS business expectations. Figure 2.10 illustrates the changes of the distribution of the VAS assessments of the present business situation over the observed time period. During the early economic upswing in 2005 the distribution of the VAS business situation is not significantly skewed. The skewness becomes significantly negative during the economic expansion, in the second half of 2006 and remains significantly negative during 2007. In 2008, as German economy cools, the shape of the distribution again changes, towards a more symmetrical pattern. Figure 2.11 shows the changes of the skewness of the VAS based business expectations and the VAS assessments of the present business situation. When the distribution of the VAS business expectations is significantly negatively skewed, as it was the case in 2005, there are more data in the left tail and at the negative side of the VAS scale than would be expected in a normal or symmetrical distribution (see Annex, Figure 2.12). In the case of the negatively skewed distribution of the VAS business

situation, there are also more data in the left tail, but the overall distribution is shifted to the right, to the positive side of the VAS scale (see Figure 2.12). This was particularly the case during the economic expansion, at the end of 2006 and the beginning of 2007 (see Figure 2.11). The striking difference between the two distributions is, that business expectations are significantly skewed to the negative side during the early upswing, while the distribution of the assessments of the present business situation is negatively skewed during the consolidated upturn and the cyclical peak. As business expectations lead the assessments of the present business situation the skewness pattern of the two variables is different in different periods of the business cycle.

According to the hypothesis of Doepke and Fritsche (2006), if the distribution of economic expectations is significantly skewed, consensus among forecasters is rejected. Although, conclusions from surveys of professional forecasters are not fully transferable to business surveys, one would expect some similarities, as business expectations also contain a macroeconomic forecast. The empirical results of the present study, however, do not support the interpretation of non-skewness as consensus. The findings rather suggest that consensus shifts over the business cycle, from the negative side of the scale to the positive and vice versa. Normality or a symmetrical distribution of business expectations appears only in particular periods, in all probability only during cyclical turning points: periods around a cyclical peak or a trough. As longer time-series of VAS data become available more explicit conclusions may be drawn.

4. Discussion

The article introduced a VAS-based measurement method of business expectations and proved the validity of various derived indicators.

Predictive validity was established by comparing the business expectations with the assessment of the business situation in subsequent surveys. VAS business

expectations were found to be a better predictor of future realizations, than business expectations elicited with the three-category scale.

In line with past studies, the concurrent validity was established by evaluating whether the VAS-based BCI exhibits the same direction and magnitude of correlation with the economic growth in the German manufacturing sector as the three-category based BCI measure of business confidence. The VAS-based business confidence indicator was found to be similarly strongly correlated with the production index growth in the German manufacturing sector as the three-category based indicator.

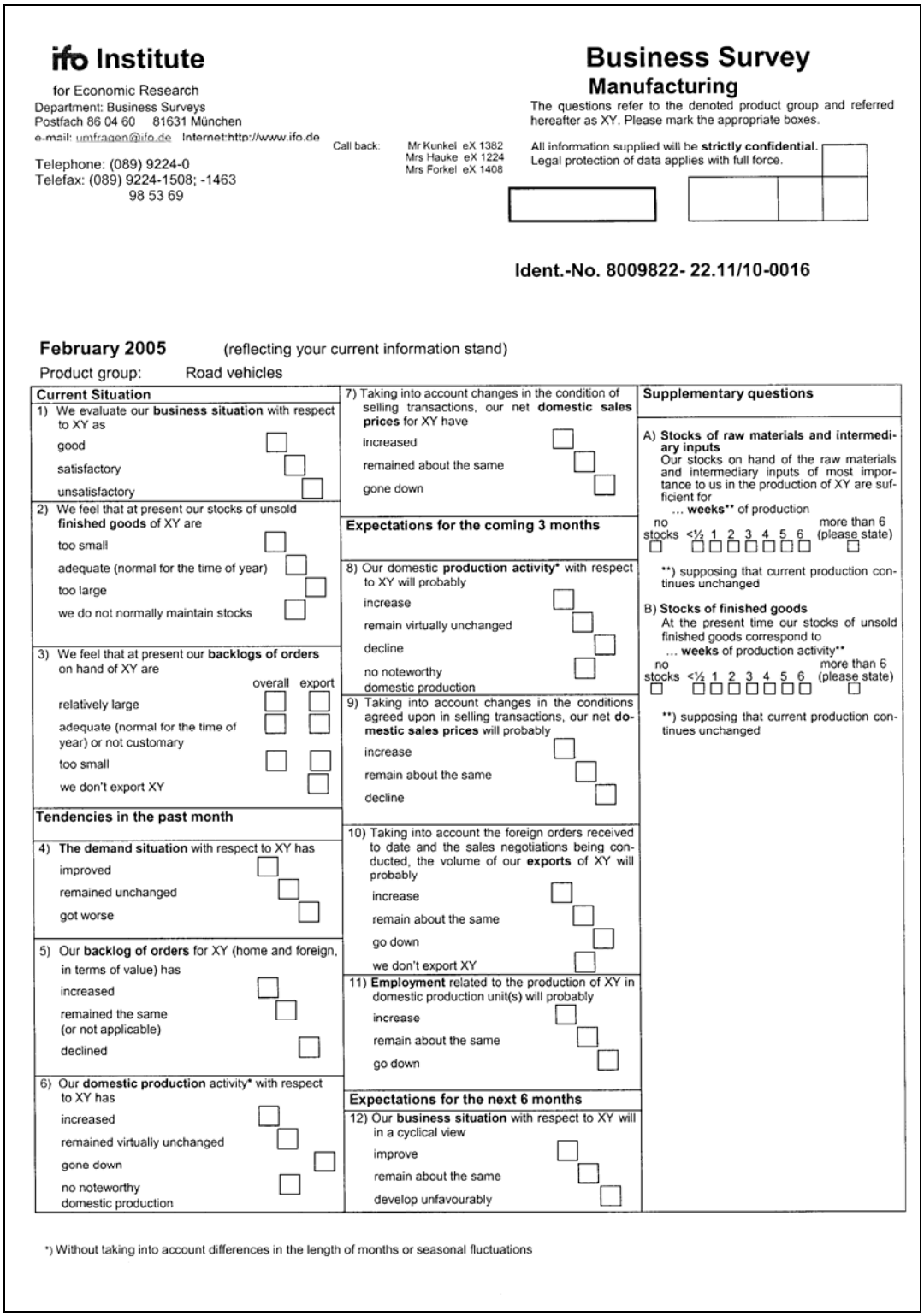
The effectiveness of the VAS as a measurement instrument of business expectations was explored not only by deriving direct measures of business confidence. As VAS data allow the estimation of higher moments of distribution, several measures of dispersion were tested for the presence of a cyclical pattern: the standard deviation, quasi-standard deviation, inter-quartile range, kurtosis and skewness. Furthermore, VAS data were found to contain information on individual-specific uncertainty about the future economic development. The dispersion measures of business expectations and the "epistemic" uncertainty measure were found to be closely related to each other. Although, the dispersion measures of business expectations appear to contain two components: uncertainty about the path of the macro-economic development and heterogeneity due to firm-specific characteristics.

The results indicate that there is an empirical relationship between the rate of change of the industrial production index and the standard deviation and the kurtosis of business expectations as well as the derived uncertainty measure. These findings are in line with the earlier findings reported in the relevant literature (Ramey and Ramey, 1995, Sepulveda-Umanzor, 2003, Mitchell et al., 2005, Doepke and Fritsche, 2006): Economic uncertainty and heterogeneity of business expectations were found to be negatively related to the economic growth. Skewness of business expectations appears to be procyclical, with a skewness close to zero pointing to a cyclical turning-point. In contrast, the skewness of the VAS based measure of the present business situation appears to be contra-cyclical.

Although the number of 34 months is by no means large, it seems reasonable to draw some tentative conclusions and to regard the findings as a "roadmap" for further research. VAS appears to be not only a valid but also a highly efficient measurement instrument of business expectations: First, it is easy to apply and does not require any quantitative information from the respondents. It consequently does not increase the cognitive load on the respondents, while the information collected is close to an interval scale measurement. Second, VAS delivers the direct measure of business expectations. It consequently supersedes the assumptions about indifference thresholds of the three-category based measurement. Third, VAS delivers valuable information on the dispersion and the skewness of business expectations, providing a more comprehensive picture of the businesses' present state and expectations. Fourth, VAS dispersion measures also contain information on the heterogeneity of expectations and macroeconomic uncertainty. Although the time-span is too short to draw conclusions about the value of these measures for forecasting, the results show, that they explain the rate of change of the industrial production index to a considerable degree, even in the very short time-period. And last but not least, VAS enables the extraction of information on uncertainty that is contained in neutral responses, without eliciting quantitative forecasts or probabilistic information.

Annex Chapter 2

Figure 2.1. The Questionnaire in the Manufacturing Sector Survey (Translated from German into English Maintaining the Original Format)



ifo Institute

for Economic Research
Department: Business Surveys
Postfach 86 04 60 81631 München
e-mail: umfragen@ifo.de Internet:http://www.ifo.de

Telephone: (089) 9224-0
Telefax: (089) 9224-1508; -1463
98 53 69

Call back: Mr Kunkel eX 1382
Mrs Hauke eX 1224
Mrs Forkel eX 1408

Business Survey
Manufacturing

The questions refer to the denoted product group and referred hereafter as XY. Please mark the appropriate boxes.

All information supplied will be **strictly confidential.**
Legal protection of data applies with full force.

Ident.-No. 8009822- 22.11/10-0016

February 2005 (reflecting your current information stand)

Product group: Road vehicles

Current Situation

1) We evaluate our **business situation** with respect to XY as
- good
- satisfactory
- unsatisfactory

2) We feel that at present our stocks of unsold **finished goods** of XY are
- too small
- adequate (normal for the time of year)
- too large
- we do not normally maintain stocks

3) We feel that at present our **backlogs of orders** on hand of XY are

	overall	export
relatively large		
adequate (normal for the time of year) or not customary		
too small		
we don't export XY		

Tendencies in the past month

4) **The demand situation** with respect to XY has
- improved
- remained unchanged
- got worse

5) Our **backlog of orders** for XY (home and foreign, in terms of value) has
- increased
- remained the same (or not applicable)
- declined

6) Our **domestic production** activity* with respect to XY has
- increased
- remained virtually unchanged
- gone down
- no noteworthy domestic production

7) Taking into account changes in the condition of selling transactions, our net **domestic sales prices** for XY have
- increased
- remained about the same
- gone down

Expectations for the coming 3 months

8) Our domestic **production activity*** with respect to XY will probably
- increase
- remain virtually unchanged
- decline
- no noteworthy domestic production

9) Taking into account changes in the conditions agreed upon in selling transactions, our net **domestic sales prices** will probably
- increase
- remain about the same
- decline

10) Taking into account the foreign orders received to date and the sales negotiations being conducted, the volume of our **exports** of XY will probably
- increase
- remain about the same
- go down
- we don't export XY

11) **Employment** related to the production of XY in domestic production unit(s) will probably
- increase
- remain about the same
- go down

Expectations for the next 6 months

12) Our **business situation** with respect to XY will in a cyclical view
- improve
- remain about the same
- develop unfavourably

Supplementary questions

A) **Stocks of raw materials and intermediary inputs**
Our stocks on hand of the raw materials and intermediary inputs of most importance to us in the production of XY are sufficient for
... **weeks**** of production

no stocks	<½	1	2	3	4	5	6	more than 6 (please state)
☐	☐	☐	☐	☐	☐	☐	☐	☐

**) supposing that current production continues unchanged

B) **Stocks of finished goods**
At the present time our stocks of unsold finished goods correspond to
... **weeks** of production activity**

no stocks	<½	1	2	3	4	5	6	more than 6 (please state)
☐	☐	☐	☐	☐	☐	☐	☐	☐

**) supposing that current production continues unchanged

*) Without taking into account differences in the length of months or seasonal fluctuations

Table 2.1. Active Enterprises in the BTS (12/2006) and the Total of Active Enterprises by 31.12.2006 in the Manufacturing Sector in Germany by Size Classes of Employees

Data source	Size classes of employees				Total
	0–9	10–49	50–249	250<	
BTS in Manufacturing (December 20006 Internet survey)	13 1%	227 16%	572 42%	566 41%	1,378 100%
Enterprises active 31.12. 2006 in Germany according to the official statistics [1)]	215,637 77%	46,579 17%	14,834 5%	3,865 1%	280,915 100%
Ifo BTS Internet sample share in the total of active manufacturing enterprises	**0.01%**	**0.49%**	**3.86%**	**14.64%**	**0.5%**

1) Active enterprises 31.12. 2006 with taxable turnover and/or with employees liable to pay social insurance contributions in 2004. Source: German Statistical Office.

Figure 2.2. Screenshot of the Visual Analog Scale as Applied in the Internet BTS in the Manufacturing Sector (German Version)

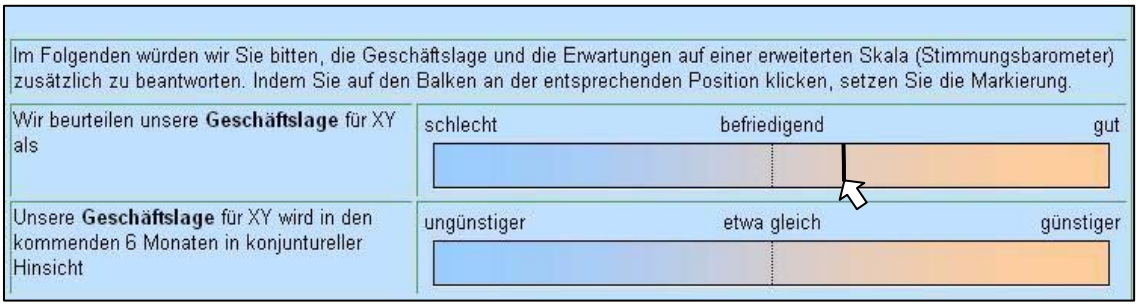
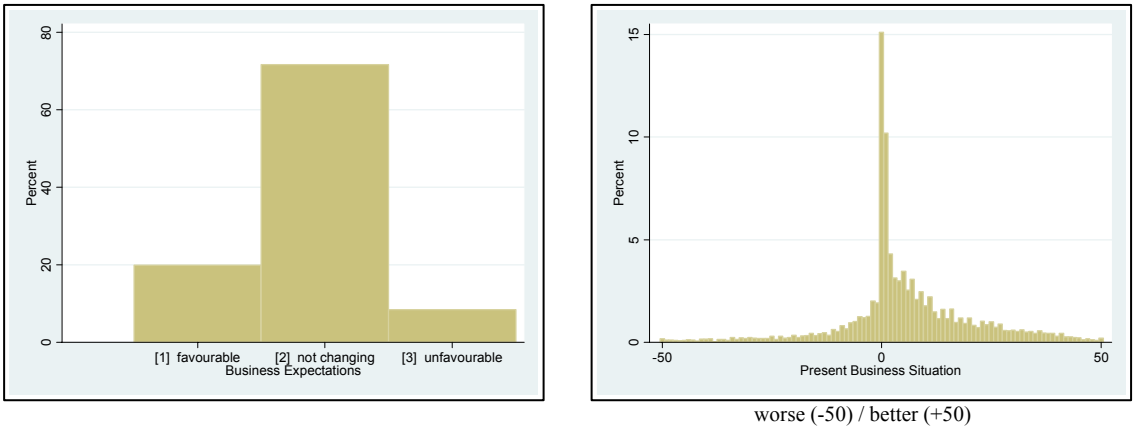


Figure 2.3. Distribution of Business Expectations on the Three-category Scale and the VAS (pooled panel data)



worse (-50) / better (+50)

Figure 2.4. Ifo Business Climate Index (seasonally adjusted) and Production Index for the German Manufacturing Sector

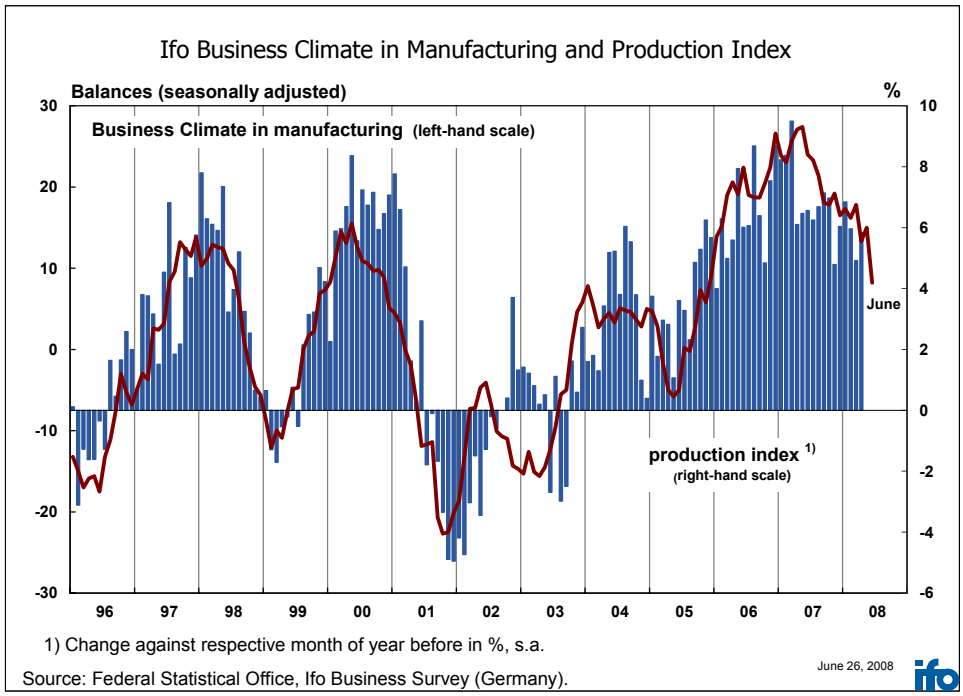


Figure 2.5. Business Climate Index (BCI) and Business Climate Index from VAS Responses (BCI-VAS) for the German Manufacturing Sector

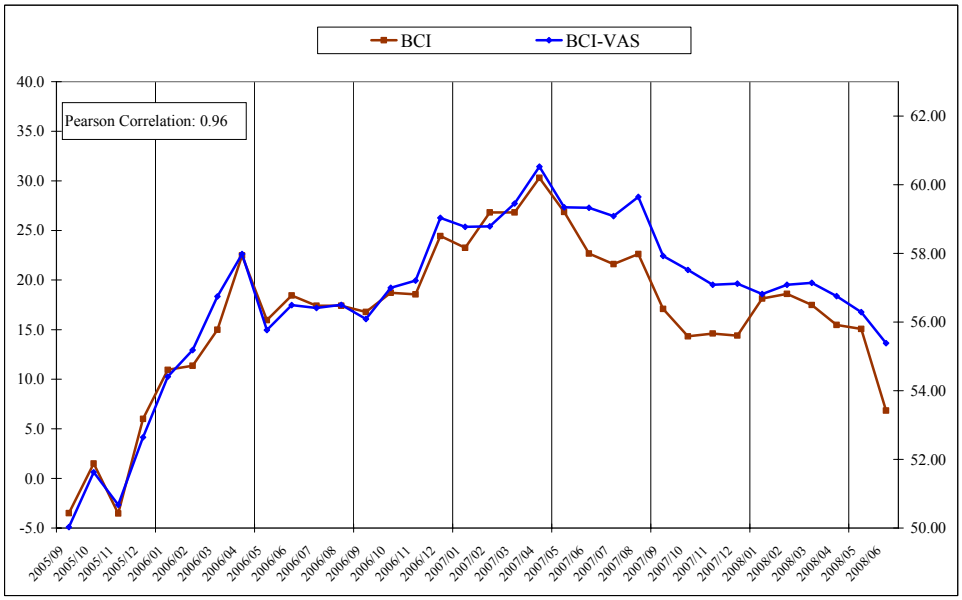


Table 2.2. Time-Series of the Production Index, Business Confidence Indicators, Dispersion Measures and "Epistemic" Uncertainty of the VAS Business Expectations

Month	PI (nsa)	PI (sa)	BCI	BCI-VAS	Standard deviation	Kurtosis	"Epistemic" Uncertainty	Quasi-standard deviation	Inter-quartile range	Skewness
2005/09	4.75	4.68	-3.51	50.03	18.12	3.48	0.18	13.00	19.00	-0.34
2005/10	5.28	5.46	1.50	51.62	17.49	3.67	0.17	12.50	18.00	-0.22
2005/11	6.24	6.24	-3.53	50.67	17.31	3.80	0.15	13.00	17.00	-0.30
2005/12	5.78	5.39	6.00	52.64	17.08	4.36	0.15	10.00	13.00	-0.41
2006/01	4.00	4.22	10.94	54.41	16.24	4.15	0.13	12.00	15.00	-0.38
2006/02	6.39	6.40	11.37	55.18	16.08	4.23	0.13	11.50	14.00	-0.32
2006/03	4.99	4.97	14.99	56.75	16.17	4.23	0.14	12.00	14.00	-0.40
2006/04	5.79	5.44	22.49	57.98	16.08	4.25	0.14	11.00	15.00	-0.31
2006/05	8.02	8.09	15.96	55.77	16.43	4.35	0.14	12.00	16.00	-0.52
2006/06	6.02	5.94	18.45	56.49	16.32	4.31	0.14	11.50	13.00	-0.34
2006/07	6.06	6.06	17.40	56.41	16.60	4.15	0.13	12.00	14.00	-0.31
2006/08	8.58	8.50	17.40	56.50	16.62	4.20	0.14	11.50	16.00	-0.31
2006/09	6.39	6.40	16.77	56.09	17.02	4.03	0.15	12.00	16.50	-0.37
2006/10	4.74	4.92	18.72	57.00	15.96	3.97	0.14	11.50	16.00	-0.29
2006/11	7.62	7.61	18.55	57.20	16.62	3.88	0.16	12.50	16.00	-0.22
2006/12	8.70	8.46	24.43	59.03	15.71	4.31	0.12	11.50	14.00	-0.18
2007/01	8.25	8.39	23.25	58.77	15.64	3.99	0.13	12.00	17.00	-0.13
2007/02	8.19	8.49	26.82	58.79	15.67	4.15	0.12	12.00	17.00	-0.21
2007/03	9.41	9.38	26.82	59.45	16.08	4.01	0.14	12.00	18.00	-0.18
2007/04	6.01	6.06	30.30	60.53	16.79	3.88	0.13	12.50	17.00	-0.17
2007/05	6.63	6.76	26.86	59.34	16.34	4.06	0.15	12.00	17.00	-0.17
2007/06	6.56	6.64	22.68	59.32	16.25	3.89	0.17	12.00	18.00	-0.15
2007/07	6.24	6.31	21.62	59.09	16.82	3.89	0.14	13.00	18.00	-0.17
2007/08	6.69	6.60	22.62	59.65	16.78	4.13	0.16	12.50	16.00	-0.23
2007/09	7.14	7.00	17.08	57.93	16.78	3.99	0.14	11.00	14.00	-0.15
2007/10	7.14	7.00	14.33	57.52	16.88	3.71	0.14	12.50	16.00	-0.03
2007/11	4.69	4.90	14.62	57.09	16.70	4.01	0.16	11.00	15.00	-0.17
2007/12	5.88	5.89	14.40	57.12	17.12	4.08	0.14	12.00	16.00	-0.17
2008/01	6.10	6.01	18.14	56.82	16.99	3.87	0.14	13.00	17.00	-0.16
2008/02	5.30	5.28	18.61	57.08	16.59	3.83	0.15	12.00	17.00	-0.11
2008/03	4.42	4.48	17.48	57.14	16.66	3.85	0.15	12.50	16.00	-0.15
2008/04	5.67	5.46	15.48	56.75	16.84	3.77	0.15	12.50	16.00	-0.14
2008/05	1.41	1.35	15.08	56.29	17.38	3.69	0.17	12.00	16.00	-0.17
2008/06	2.25	2.18	6.84	55.38	17.39	3.45	0.14	13.00	17.00	-0.11

Table 2.3. Correlation Matrix of Various Indicators of the Economic Performance in the German Manufacturing Sector

Indicators	PI(sa) [a]	PI(nsa) [b]	BCI	BCI-VAS
PI(sa)	1			
PI(nsa)	0.99*	1		
BCI(sa)	0.45*	0.44*	1	
BCI-VAS(nsa)	0.38*	0.37*	0.96*	1

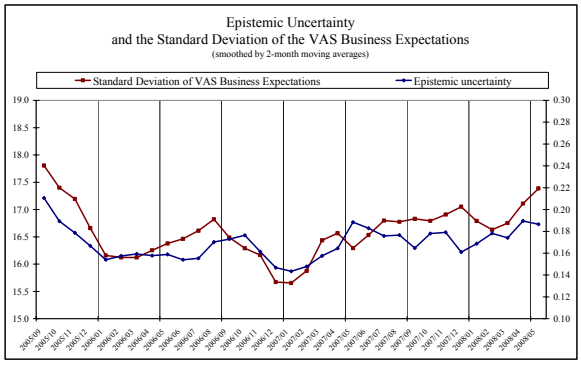
* Significance level <0.05. a) Production index (2000=100) for the manufacturing industry, value adjusted for working day and seasonal variations (X-12-ARIMA). b) Production index (2000=100) for the manufacturing industry, value adjusted for working day variations, not seasonally adjusted (X-12-ARIMA). Source: Federal Statistical Office Germany 2008.

Table 2.4. Correlation Matrix of the Dispersion Measures and the Uncertainty Proxy (9/2005 - 6/2008)

	Epist. Uncert.	Dispers. of bus. expect.	Kurt.	Quasi-standard dev. [a]	Inter-quartile range
Epistemic uncertainty	1				
Dispersion of business expect.	0.66*	1			
Kurtosis	-0.54*	-0.64*	1		
Quasi-standard deviation[a]	0.13	0.40*	-0.68*	1	
Inter-quartile range	0.37*	0.31	-0.68*	0.69*	1
Skewness	0.07	0.04	-0.56*	0.31	0.37*
1. Heterogeneity I [b]	0.00	0.76*	-0.39*	0.41*	0.09
2. Heterogeneity II [c]	0.00	-0.35*	0.85*	-0.72*	-0.57*
3. Heterogeneity III [d]	0.00	0.08	-0.52*	0.69*	0.93*

* Significance level <0.05. [a] Equals to half the difference between the 16th and 84th percentile of the sample of point forecasts. [b] Calculated as residual from the regression of the standard deviation of business expectations on "epistemic" uncertainty. [c] Calculated as residual from the regression of the kurtosis of business expectations on "epistemic" uncertainty. [d] Calculated as residual from the regression of the inter-quartile range of business expectations on "epistemic" uncertainty.

Figure 2.6. Relationship Between the Dispersion Measures of VAS Business Expectations and the "Epistemic" Uncertainty



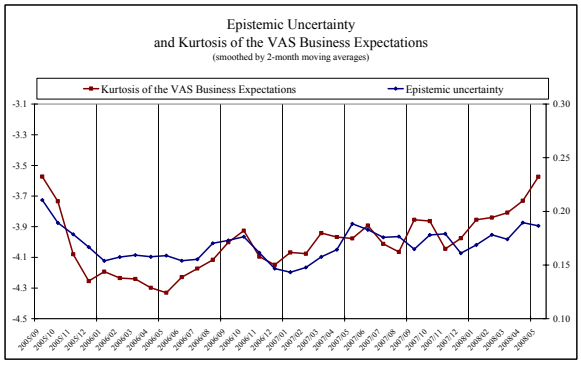


Table 2.5. Results of the Panel Fixed Effects Regression of the VAS Present Business Situation on Lagged VAS Expectations (Individual and Time Fixed Effects [a], Balanced Panel)

Regressions for Different Lags	Business Expectations Coef.	Company size [b]	R^2 [c]
Lag 1	0.39*** (0.02)	-0.72*	0.74
Lag 2	0.37*** (0.02)	-0.62*	0.74
Lag 3	0.31*** (0.02)	-0.65*	0.74
Lag 4	0.25*** (0.02)	-0.63*	0.74
Lag 5	0.21*** (0.02)	-0.62	0.75
Lag 6	0.17*** (0.02)	-0.56	0.76
Lag 7	0.15*** (0.02)	-0.54	0.76
Lag 8	0.08*** (0.02)	-0.54	0.77
Lag 9	0.06*** (0.02)	-0.71*	0.77

Significance level: * <0.10; ** <0.05; *** <0.01. Standard errors in parenthesis. a) Time-fixed effects: indicator dummy variables for each month. Each line represents a separate regression of the assessment of the present business situation on lagged business expectations, company size class and further controls (time and individual dummies). b) Company size classes: 1/49 = 1, 50/99 = 2, 100/149 = 3, 150/199 = 4, 200/249 = 5, 250/349 = 6, 350/499 = 7, 500/649 = 8, 650/999 = 9, 1000/1999 = 10, 2000/200000 = 11. c) R^2 from fixed effects OLS regression with 164 indicator dummy variables for each observation (using command *areg* in Stata).

Table 2.6. Spearman's Correlation between the Assessment of the Present Business Situation and Lagged Business Expectations (Balanced Panel)

Lagged Business Expectations	VAS	3-cat.
Lag 1	0.60	0.25
Lag 2	0.59	0.26
Lag 3	0.58	0.25
Lag 4	0.57	0.24
Lag 5	0.55	0.23
Lag 6	0.55	0.23
Lag 7	0.54	0.23
Lag 8	0.53	0.22
Lag 9	0.53	0.21

Each line shows the Spearman's rho of the correlation between the assessment of the present business situation and lagged business expectations. As N=5,576 all correlations are significant.

Table 2.7. Spearman's Correlation between the Time-series of Aggregated Business Situation and Aggregated Business Expectations For Different Lags

Lagged Business Expectations	Unweighted time-series		Weighted time-series	
	VAS	3-Cat.	VAS	3-Cat.
Lag 1	0.65*	0.26	0.56*	0.17
Lag 2	0.60*	0.38*	0.57*	0.22
Lag 3	0.58*	0.34	0.53*	0.21
Lag 4	0.52*	0.33	0.47*	0.15
Lag 5	0.49*	0.29	0.41*	0.13
Lag 6	0.46*	0.13	0.41*	0.26
Lag 7	0.37	0.14	0.32	0.29
Lag 8	0.23	0.21	0.25	0.37
Lag 9	0.03	0.32	0.03	0.39

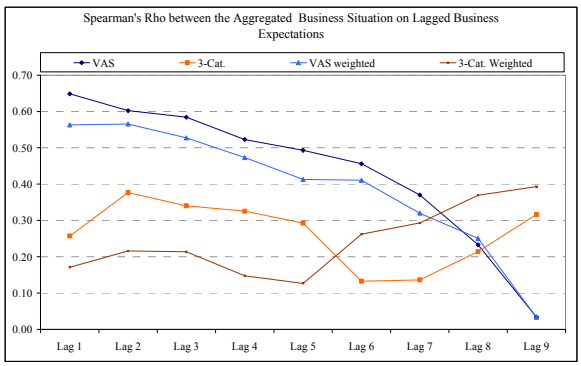
* Significance level <0.05. Each line shows the Spearman's rho of the assessment of the present business situation and lagged business expectations.

Table 2.8. R^2 of the Regression of the Time-series of the Aggregated Business Situation on Lagged Aggregated Business Expectations

Lagged Business Expectations	Unweighted time-series		Weighted time-series	
	VAS	3-Cat.	VAS	3-Cat.
Lag 1	0.66	0.08	0.51	0.04
Lag 2	0.66	0.14	0.55	0.10
Lag 3	0.63	0.15	0.50	0.13
Lag 4	0.50	0.13	0.39	0.10
Lag 5	0.43	0.07	0.30	0.07
Lag 6	0.39	0.01	0.26	0.15
Lag 7	0.28	0.00	0.19	0.17
Lag 8	0.22	0.09	0.23	0.32
Lag 9	0.08	0.21	0.09	0.33

Each line represents the R^2 from a separate regression of the present business situation time-series on the time-series of lagged business expectations.

Figure 2.7. Figures to the Table 2.7 and Table 2.8



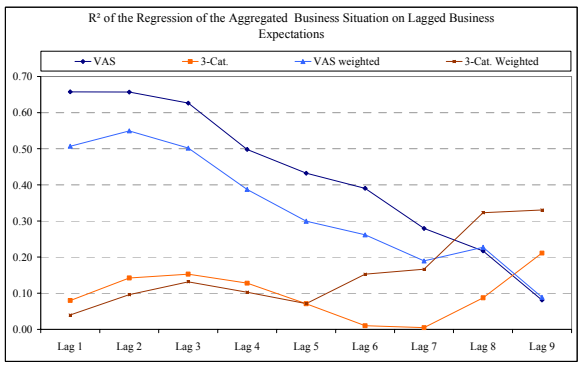
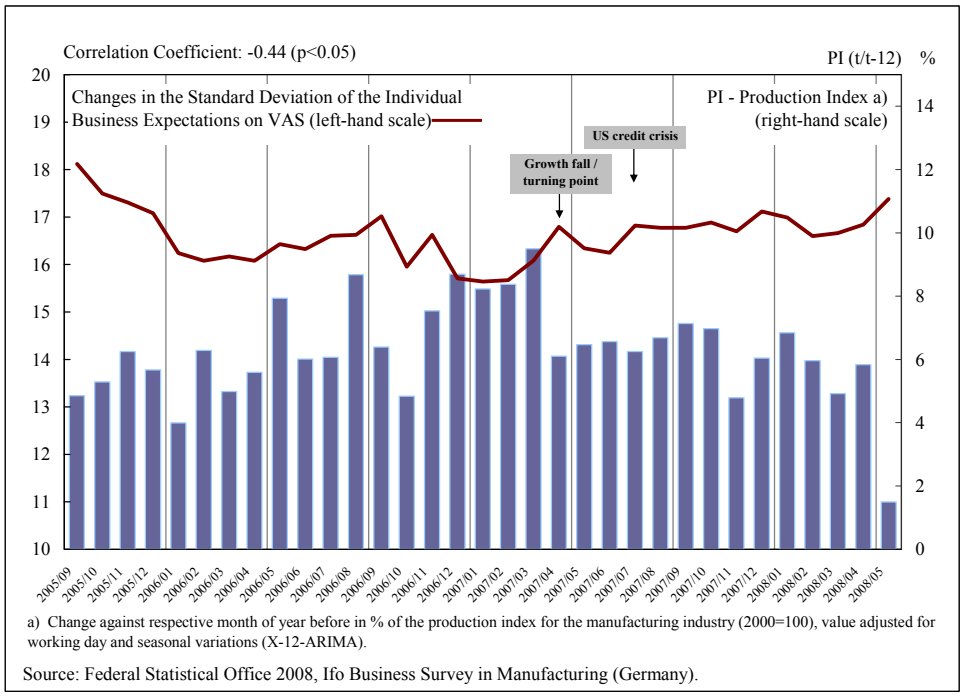


Figure 2.8. Dispersion of Business Expectations and Production Index(sa) in the German Manufacturing Sector



a) Change against respective month of year before in % of the production index for the manufacturing industry (2000=100), value adjusted for working day and seasonal variations (X-12-ARIMA).

Source: Federal Statistical Office 2008, Ifo Business Survey in Manufacturing (Germany).

Table 2.9. Correlation Matrix of the Dispersion Measures, "Epistemic" Uncertainty and Heterogeneity of Business Expectations with Various Indicators of Economic Growth in the German Manufacturing Sector

Indicators	PI(nsa)[a)]	PI(sa)[b)]	BCI	BCI-VAS
Epistemic uncertainty	-0.44 *	-0.44 *	-0.47 *	-0.41 *
Dispersion of business expect.	-0.50 *	-0.52 *	-0.69 *	-0.61 *
Kurtosis	0.47 *	0.45 *	0.38 *	0.26
Quasi-standard deviation [c)]	-0.15	-0.12	-0.16	-0.09
Inter-quartile range	0.02	0.06	-0.02	-0.01
Skewness	-0.03	-0.02	0.30	0.46 *
Heterogeneity I [d)]	-0.28	-0.30	-0.51 *	-0.45 *
Heterogeneity II [e)]	0.27	0.26	0.15	0.05
Heterogeneity III [f)]	0.20	0.24	0.16	0.15

* Significance level <0.05. a) Production index (2000=100) for the manufacturing industry, value adjusted for working day variations, not seasonally adjusted (X-12-ARIMA). b) Production index (2000=100) for the manufacturing industry, value adjusted for working day and seasonal variations (X-12-ARIMA). Source: Federal Statistical Office Germany 2008. c) Equals to half the difference between the 16th and 84th percentile of the sample of point forecasts. d) Calculated as residual from the regression of the standard deviation of business expectations on "epistemic" uncertainty. e) Calculated as residual from the regression of the kurtosis of business expectations on "epistemic" uncertainty. f) Calculated as residual from the regression of the inter-quartile range of business expectations on "epistemic" uncertainty.

Figure 2.9. Distribution of Business Expectations in the German Manufacturing Sector Over Time

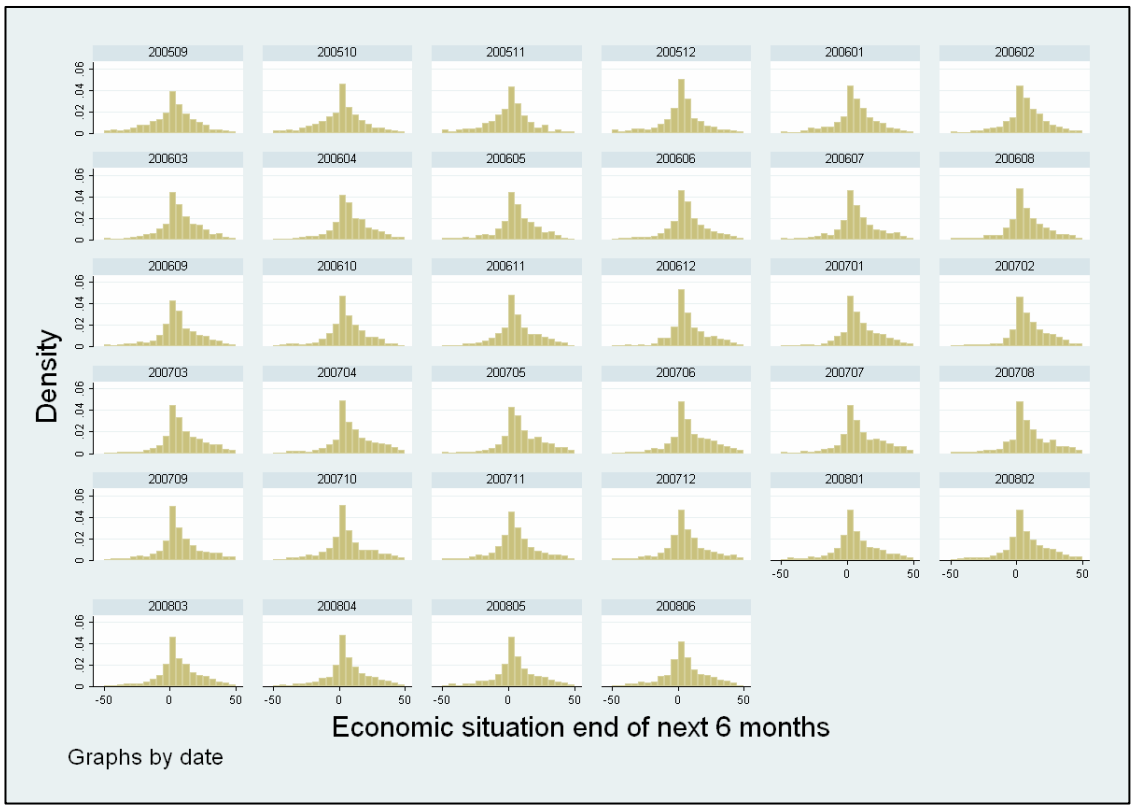


Table 2.10. Correlation Matrix of Further VAS Indicators and Indicators of Economic Growth in the German Manufacturing Sector

	PI(nsa)[a)]	PI(sa)[b)]	BCI	BCI-VAS
Neutral responses in VAS bus. situation	-0.37 *	-0.35 *	-0.28	-0.28
Dispersion of the VAS bus. situation	-0.26	-0.27	-0.75 *	-0.74 *
Kurtosis of the VAS bus. situation	0.43 *	0.43 *	0.72 *	0.66 *
Inter-quart. range of the VAS bus. sit.	-0.12	-0.12	-0.29	-0.23
Skewness of the VAS bus. Situation	-0.49 *	-0.51 *	-0.81 *	-0.75 *

* Significance level <0.05. a) Production index (2000=100) for the manufacturing industry, value adjusted for working day variations, not seasonally adjusted (X-12-ARIMA). b) Production index (2000=100) for the manufacturing industry, value adjusted for working day and seasonal variations (X-12-ARIMA). Source: Federal Statistical Office Germany 2008.

Figure 2.10. Distribution of the Assessments of the Present Business Situation in the German Manufacturing Sector Over Time

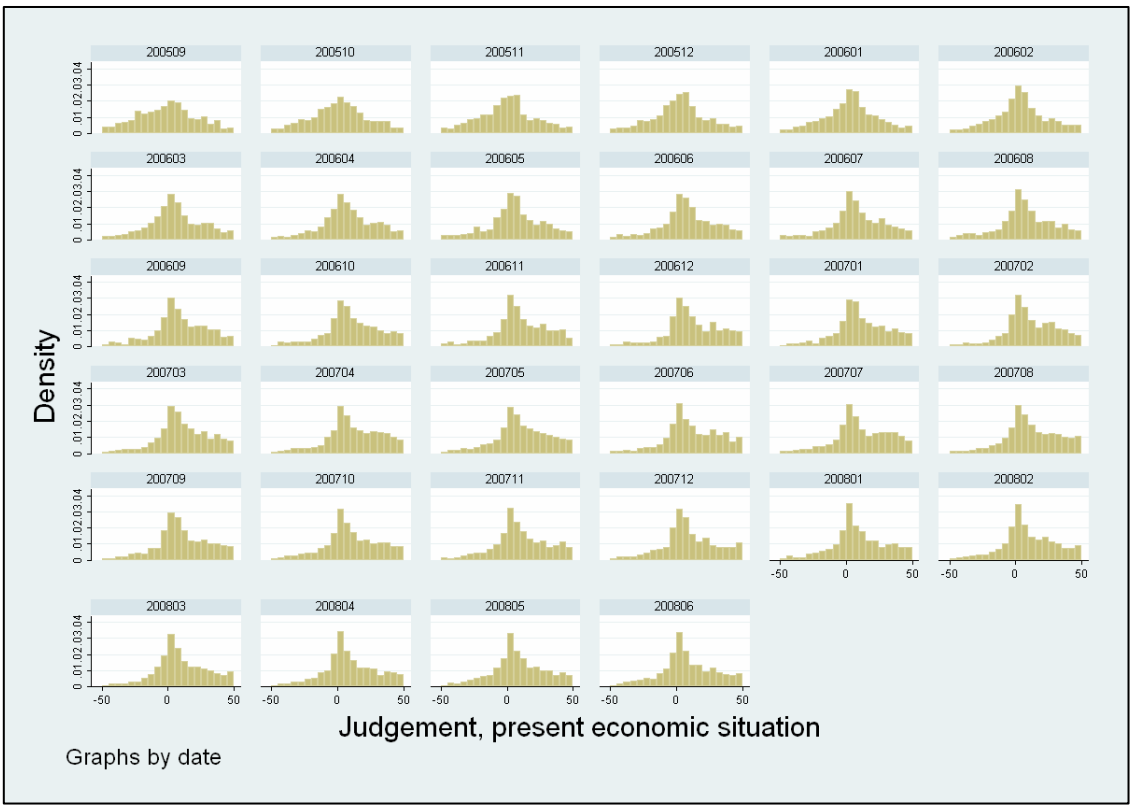


Figure 2.11. Skewness of the Distribution of the VAS Business Expectations and of the VAS Assessment of the Present Business Situation over Time

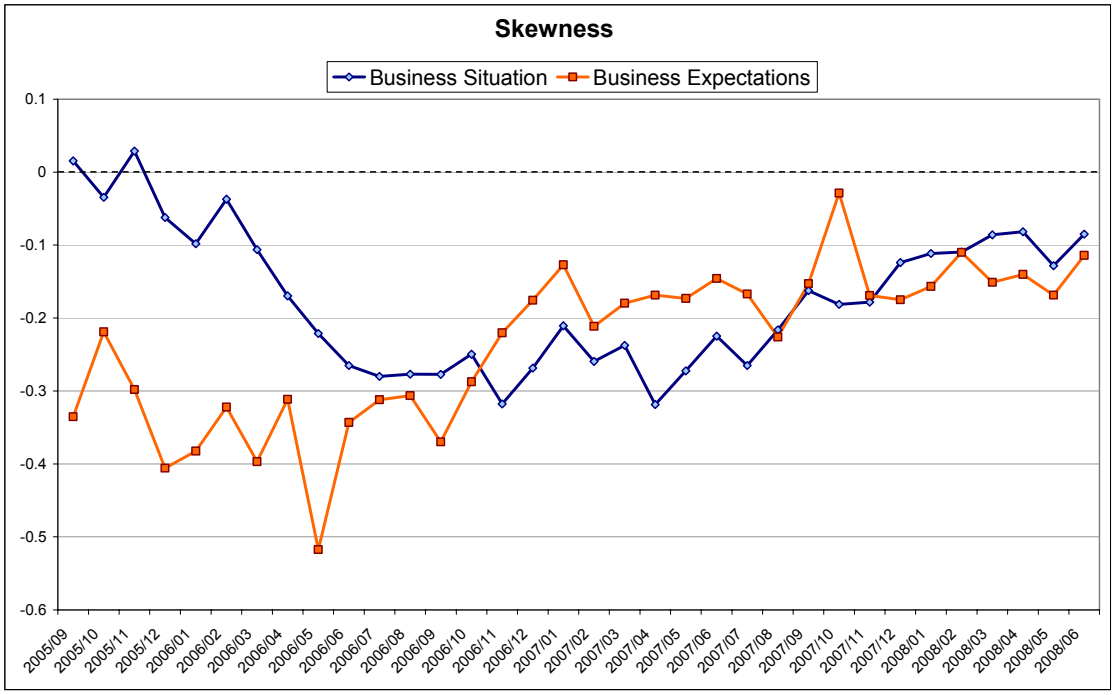
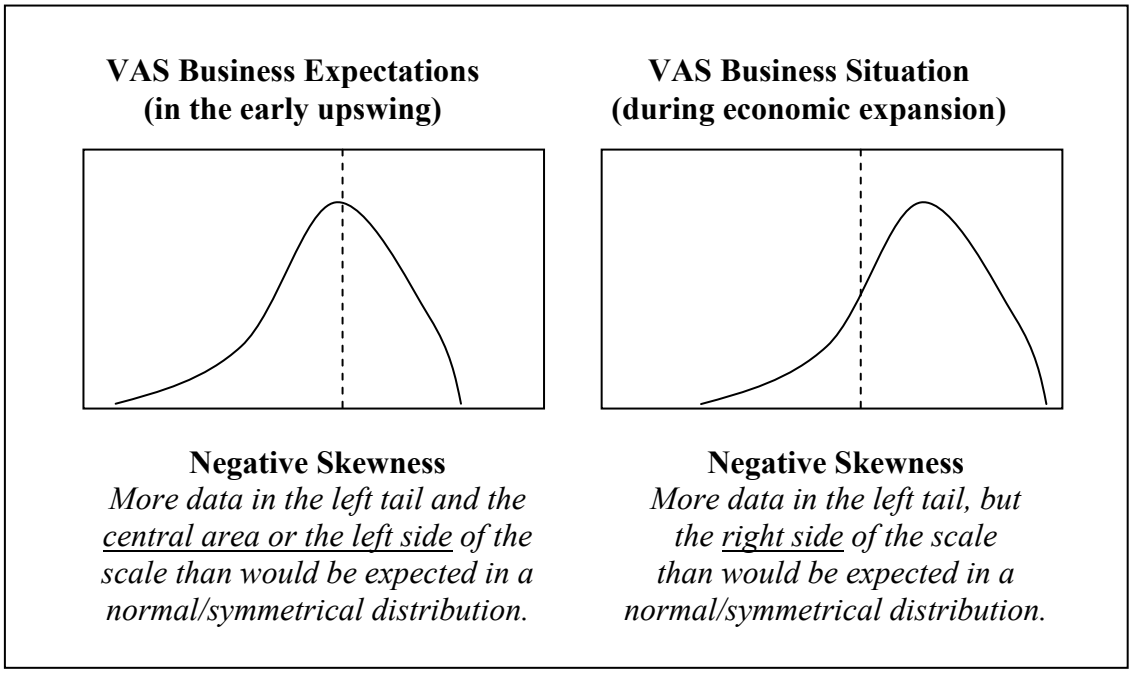


Figure 2.12. Prototypical Pattern of the Skewness of the VAS Business Expectations and the Skewness of the VAS Business Situation



Chapter 3

Testing the Assumptions of Three-category Based Business Expectations Using the Visual Analog Scale

Abstract

The modeling of responses to a category-rating scale implies a latent variable model with indifference thresholds marking the values of a latent variable at which a respondent is indifferent between categories. Several quantification methods of qualitative business expectations, amongst them the balance statistics (fraction of positive responses minus the fraction of negative responses), rely on strong assumptions relating the indifference thresholds, postulating that the indifference thresholds are symmetrical and neither vary over time nor across individuals. The aim of the study is to test whether these assumptions hold for business expectations measured with the three-category scale. Data elicited with the visual analog scale, which deliver the direct measurement of business expectations, are used to test these assumptions.

1. Introduction and Background

Business tendency survey (BTS) data are widely used to track and to forecast macroeconomic variables. BTS as a rule collect qualitative information on the variables of interest. The motivation and popularity of using qualitative variables in BTS roots in the fact that it is generally much easier and faster to collect qualitative rather than quantitative information (OECD 2003). The measurement procedure involves three general categories: positive replies, indifferent replies and negative replies. The experience of interviews and mail surveys suggested the usage of category-rating scales with three or at most five categories (very good/good, satisfactory, bad/very bad). However, the category responses, although broadly applied and pursued by the European Commission,[1] are coarse proxies of agents' true expectations (Nardo, 2005). In order to extract quantitative indicators, quantification methods are applied to categorical business survey responses relating them to aggregate realizations. Most quantification methods of qualitative responses are based on assumptions, which if violated can introduce a systematic bias in resulting indicators.

Balance statistics (fraction of positive responses minus the fraction of negative responses) is the most popular quantification method of business expectations applied in practice and used by Eurostat to track the official data on economic growth, such as industrial production. The balance statistics of responses to a three-category scale implies a latent variable model with indifference thresholds marking the values of the latent variable at which a respondent is indifferent between categories. The balance statistics is based on the assumption that this indifference interval[2] is symmetric. Applying these assumptions to business expectations about the company's output growth, the balance statistics provides an accurate measure of

[1] European Commission Directorate-Generale for Economic and Financial Affairs website contains detailed information on the joint harmonized EU program of business and consumer surveys: http://ec.europa.eu/economy_finance/indicators/businessandconsumersurveys_en.htm (retrieved 06/13/2007).

[2] The term indifference interval was used by Theil (1952). Carlson and Parkin refer to it as to "difference limen". Batchelor (1986) uses a term from the signal detection theory "just noticeable difference" (j.n.d.) and Dasgupta and Lahiri (1992) calls the indifference thresholds "imperceptibility parameters".

average output growth if the percentage change in output making firms report a fall and the percentage change in output making firms report a rise are the same. This quantification method relies on further strong assumptions: It assumes that the thresholds do not vary over time and are identical across the respondents.

The theoretical foundations of the balance statistics have been laid by Anderson in 1952. Alternatively two streams of quantification methods evolved over time, the probability method motivated by Theil (1952, 1955) and rediscovered by Carlson and Parkin (1975) and the regression method developed by Pesaran (1984). Later, Cunningham, Smith and Weale (1998) offered an alternative reverse-regression approach, while Seitz (1988), Löffler (1999), Fishe and Lahiri (1981) and Batchelor and Orr (1988) contributed to the refinement of the Carlson-Parkin method. The three assumptions, that the threshold interval is symmetrical, neither varies over time nor across companies, apply to the balance statistics and to the original Carlson and Parkin quantification method. While the regression-based quantification methods of Pesaran and other advocate the idea of asymmetrical responses, especially for particular time periods, for example periods of rising inflation (Pesaran, 1984, Batchelor, 1986). Detailed discussions on quantification methods can be found by Nardo (2003) and Pesaran and Weale (2005).

Theoretical considerations and empirical findings suggest that the threshold interval may be asymmetrical, vary over time and also differ across respondents (Seitz, 1988, Pesaran, 1984, Dasgupta and Lahiri, 1992, Batchelor and Orr, 1988, Ronning, 1990). For example, for larger companies a smaller percentage changes in output may result in reporting a rise or a fall, than for small-scale enterprises (Ronning, 1990). Adequately, during economic expansion (downswing) a higher (lower) percentage change in output is necessary to make firms reporting a rise (or a fall). The relevant literature that evaluates the symmetry and constancy of indifference thresholds can be divided in two streams, according to the empirical material of the studies: quantification of inflation and output expectations from surveys and studies using simulation experiments.

Relatively few authors investigated qualitative output expectations. Bennett (1984) applied the Carlson-Parkin quantification method on the industrial trends survey data for United Kingdom to construct time-series of output expectations for the manufacturing industry. He criticizes the assumption on the constancy of indifference thresholds over time and their independency from the magnitude of change in output, although he did not investigate this problem further. Mitchell et al. (2002b) propose a quantification method for manufacturing output growth expectations based on time-variant thresholds and find that the resulting indicator outperforms traditional indicators that assume time-invariant thresholds relating the quantification of aggregated time-series of the manufacturing output growth.

The vast majority of studies, however, dealt with inflation expectations. As the quantification methods of qualitative inflation expectations are based on the same assumptions, these studies are also briefly discussed. Some of the earliest studies were conducted by Batchelor (1985, 1986, 1988) and Pesaran (1984). The authors found no support for the assumptions of constant indifference thresholds in qualitative inflation expectations, and suggested to model the indifference thresholds as dependent variables of the inflation rate and a proxy of inflation uncertainty (measured by the dispersion of price changes across industry). In contrast Seitz (1988) did not find that the threshold parameters are dependent of the level of inflation or its dispersion. Dasgupta and Lahiri (1992) used producer price index in the manufacturing sector as a benchmark for the quantification of the US inflation expectations of the National Association of Purchasing Managers Survey. They tested for different shapes of distribution and allowed the indifference interval to be asymmetric and variable over time. They found that allowing for time-varying thresholds leads to unbiasedness of expectations, although the forecast error decreases only modestly. The authors also found a significant asymmetry in thresholds. Henzel and Wolmershäuser (2005) asked survey respondents directly for the minimum value to which inflation rate has to increase (decrease) before respondents would report an UP or DOWN of inflation. They also found time-varying and asymmetrical indifference thresholds, with the upper bound somewhat higher than the lower bound.

Nardo and Cabeza-Gutes (2005) compared in a simulation study the ability of the three major quantification methods (balance statistics, Carlso-Parkin method and the Pesaran regression approach) to trace actual variables in the presence of a measurement error. Through simulations they isolated the measurement error introduced by incorrect assumptions when quantifying survey results. They found that the Carlson-Parkin method with time-varying thresholds approximates true expectations best. Similar results were reported by Claveria et al. (2006).

Another crucial assumption in balance statistics and the original Carlson-Parkin method is the normality of the distribution of respondents' views about changes in the respective variables. It has been first claimed by Foster and Gregory (1975) that the distribution of responses cannot be assumed to be consistently symmetric. It is likely to be positively (negatively) skewed when the average expected change is positive (negative). Later other authors supported this idea (Batchelor, 1981, 1988, Dasgupta and Lahiri, 1992). Batchelor (1981) finds that using an inflation expectations distribution with systematically varying skewness can eliminate symptoms of irrationality often found in survey-based expectations.

Summarizing the literature review, the majority of studies question the normality assumption, argue for asymmetric and stochastic indifference thresholds, and consequently for quantification methods which can model this kind of data. However, in almost all empirical studies the qualitative expectations are linked to aggregate realizations. Furthermore, the fast majority of studies deals with inflation expectations. Studies dealing with business expectations that can be linked to output growth are rare. The reason is that direct reference series to business expectations are not as readily available as reference series to inflation expectations. It is particularly difficult to collect this kind of data at the micro-level, as business profits and output growth are variables which companies as a rule are not willing or able to report on. In the present study the data collected with the visual analog scale (VAS) are used as proxy for the companies' output growth, to test the three assumptions about the indifference thresholds: symmetry, constancy over time and across companies. As a by-product of this paper it is also tested whether business expectations are normally and symmetrically distributed, another assumption often made in the statistical

analysis of three-category responses. VAS offers an attractive alternative to a quantitative measurement of profits or output growth as respondents are not asked to report quantitative information, but can use the VAS to directly express the level of their business confidence. Consequently the VAS presents a direct measure of business expectations that supersedes the necessity to make assumptions about the distribution, as it can be observed directly.

The present study contributes to the existing literature in several ways. In the previous studies the data used as reference-series to the qualitative business expectations are based on official macro-economic time-series and not on individual survey responses. In the present study both – the qualitative business expectations and the reference-series – are survey-based. The data of the study are based on a panel of 34 consecutive survey months. This fact facilitates tests on the variety of the indifference thresholds over time. Furthermore, the paper also addresses the variation of indifference thresholds across individuals, an issue that to the best of our knowledge, has not been yet addressed in the literature (see also Nardo, 2003). Also, the method suggested in the present study, to collect qualitative response on a continuous scale, has not been applied in the history of business surveys.

The plan of the paper is as follows. The next section describes the data set. Section 3 is dedicated to the statistical framework and the methods applied to test the assumptions about the indifference thresholds. Section 4 contains the empirical results. The paper concludes with a discussion of results and inference on whether empirical evidence provides support for the most common assumptions underlying statistical analysis and several quantification methods of qualitative business expectations.

2. Data

The data used in the study have been collected within the monthly business survey in the German manufacturing sector. This data set has been already used in Chapter

2. This section therefore corresponds in various points to the section 2.1 of the Chapter 2.

The used business survey elicits information from companies on a number of business variables, such as the assessment of the present business situation and business expectations (a sample paper questionnaire is provided in the Annex to the Chapter 2, Figure 2.1). The panel involves business officials with specializations within their companies in management, finance, and other strategic business functions. The companies covered by the sample include small, medium-sized and large companies. The survey participation is absolutely voluntary and derives entirely from the interest in the survey results, as no other compensation is offered. Becker and Wohlrabe (2008) provide a detailed description of the Ifo Business Survey micro data.

The assessments of the present business situation and business expectations have been elicited twice within the questionnaire: with the three-category scale and the VAS (a screen shot of the VAS is provided in the Annex of the Chapter 2, Figure 2.2.).

Observations containing missing and inconsistent responses to the questions of interest were excluded from the data-set. Inconsistencies appeared when a respondent has chosen a positive category on the three-category scale but placed the marker in the negative territory of the VAS, or vice versa.

The data set of individual responses covers 34 survey waves, beginning with the September 2005 survey and continuing to the June 2008 survey. The overall data set contains 45,883 observations. The data of the study comprise a conventional incomplete panel dataset, as the sample changes from time to time and not every respondent provides an expertise every month. However, the mean number of survey participations during the 34-months period is 27, indicating that the panel is sufficiently stable. The majority of the following analyses were thus conducted using the whole data-set. Furthermore, creating a balanced panel would result in loss of

degree of freedom. The following analysis is conducted the whole data set that is summarized in Table 3.1 of the Annex.

The Figure 3.1 illustrates the distribution of responses to the three-category scale and the VAS for the two variables of interest – present business situation and business expectations. In the distribution of the VAS responses there is a peak at the central anchor of the scale. As has been discussed in Chapter 2, the central anchor is not only necessary to give respondents an orientation on the VAS and to allow for a more subtle distinction of their preferences. The neutral responses can also serve as an indicator of "epistemic uncertainty" (Bruine de Bruin et al., 2000). As neutral VAS responses were found to contain a different information than VAS responses in other areas of the scale, for the following analysis responses at the central anchor (+/- 1) were removed from the data set.

3. Statistical Framework

The modeling of responses to an ordinal scale implies a latent variable model (Finney, 1971). Though a qualitative assessment seems to be very simple, it is underpinned by a complex cognitive process of summarization (Tourangeau et al., 2000). Whether the business situation of a company can be described with "good" depends along with some objective factors (profits etc.) on the context of the questions, such as the situation of a company, in relation to the situation of other companies. Thus, business confidence is seen as a function of many underlying economic factors that are considered by the surveyed expert to influence economic improvement or deterioration (OECD 2003):

$$Y = f(x_i) = \beta_0 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + e \qquad (1)$$

where

Y – Business confidence (*attitude*)
x_i – Set of explanatory variables (*market prices, demand, competitors, macro-economic conditions, respondent's characteristics, such as pessimism or optimism etc.*)

The VAS measures the level of Y directly. In contrast, within a three-category assessment the data generating process has two stages. First stage is formally described with equation (1), where the respondent takes into account the factors which are regarded to be important. In the second stage the respondent has to trigger the continuous attitude variable into category ratings and to choose a category. Which category the respondent chooses depends not only on the value of Y but also on the respondent's indifference thresholds. When Y falls below a particular threshold a, the respondent will opt for a negative category (D). When Y overcomes b, for the positive category (U) and when Y is inside the thresholds a and b the respondent will opt for the central category (S). The observed variable is the ordinal response (*D, S* or *U*). Given the model of the latent variable Y presented in equation (1), the responses of a company can be classified as follows:

D – "down" is observed if $Y < a$
S – "same" is observed if $a \leq Y \leq b$
U – "up" is observed if $Y > b$

Assuming that Y relates to an unobserved variable Y^* (for example output growth rate) by the structural model $Y=Y^*+e$. Substituting for Y we can rewrite the above statement:

$$\Pr(U \mid Y^*) = \Pr(\infty > Y^* + \varepsilon > b)$$
$$\Pr(S \mid Y^*) = \Pr(Y^* + \varepsilon \leq b) - \Pr(Y^* + \varepsilon \leq a)$$
$$\Pr(D \mid Y^*) = \Pr(-\infty < Y^* + \varepsilon < a)$$

Subtracting Y^* within the inequality we receive:

$$\begin{aligned} \Pr(U \mid Y^*) &= \Pr(\infty - Y^* > \varepsilon > b - Y^*) \\ \Pr(S \mid Y^*) &= \Pr(\varepsilon \leq b - Y^*) - \Pr(\varepsilon \leq a - Y^*) \\ \Pr(D \mid Y^*) &= \Pr(-\infty - Y^* < \varepsilon < a - Y^*) \end{aligned} \tag{2}$$

Assuming the standard normal distribution of Y^*, we receive the following probability distribution, with the cumulative standard normal distribution function Φ.

$$\Pr(U \mid Y^*) = 1 - \Phi(\frac{b - Y^*}{\sigma})$$
$$\Pr(S \mid Y^*) = \Phi(\frac{b - Y^*}{\sigma}) - \Phi(\frac{a - Y^*}{\sigma}) \quad (3)$$
$$\Pr(D \mid Y^*) = \Phi(\frac{a - Y^*}{\sigma})$$

The probability distribution of responses to a three-category scale depends not only on the value of the unobserved continuous variable but also on the function that forms the indifference thresholds. Consequently, for the analysis and interpretation of the data, assumptions about the form of the threshold interval are necessary.

One of the crucial assumptions about the form of the indifference interval is that the interval thresholds are symmetrical ($a = -b$). This assumption is tested in the present study by the following empirical strategy.

Within a longitudinal survey, respondents switch from one category to another, when the value of the variable Y from equation (1) overcomes a particular indifference threshold. Consequently, nine transition states exist that define the nine switching regimes of the three-category scale.

Illustration of the Transition States within a Three-category Scale in Two Consecutive Survey Waves.

Month 1 / **Month 2**	**Positive UP (U)**	**Neutral SAME (S)**	**Negative DOWN (D)**
Positive / UP (U)	UU	**SU**	*DU*
Neutral / SAME (S)	**US**	SS	**DS**
Negative / DOWN (D)	*UD*	**SD**	DD

The negative indifference threshold a lies within the two transition states *SD* and *DS*, when respondents switched from the neutral category into the negative category and vice versa. The positive indifference threshold b lies within the two transition states *US* and *SU*, when respondents switch from the positive category into the neutral category and vice versa. The two transition states when respondents skipped the neutral category and switched from one extreme category in the other extreme

category (*UD* and *DU*) were seldom in the present data set (less than 1 percent) and are not considered in the analysis, assuming that this response behavior results from some positive or negative shock.

At the same time, in each month and for each variable of interest – present business situation and business expectations – *Y* is recorded on the VAS. Assuming stochastic distribution of indifference thresholds in the population, the mean VAS value over two consecutive months in the transition states *US* and *SU* represents the mean upper indifference threshold *b*. Accordingly, the mean VAS value over two consecutive months in the transition states *DS* and *SD* represents the mean lower indifference threshold *a*.

Figure 3.2 in the Annex illustrates the distribution of the indifference thresholds in the data set. The above mentioned four of nine possible transition states (positive/neutral, neutral/positive, neutral/negative, negative/neutral) are placed at the abscissa of the graph. The ordinate illustrates the corresponding distribution of the indifference thresholds on the VAS. Figure 3.2 already demonstrates that there is a high variation of indifference thresholds across respondents. The figure also illustrates that the distribution of thresholds of those respondents who switched from the positive to the neutral category and respondents who switched from the neutral to the positive category is very similar. The same applies to those respondents who switched from the negative in the neutral category or from the neutral to the negative category. Table 3.2 in the Annex contains the results of a two-tailed t-test on the equality of means and confirms that there is no significant difference in threshold means between the states SU and US and the states DS and SD in the variable "present business situation". Consequently, the two positive transition states have been combined to one transition state (positive/neutral/positive-USU) with an identical threshold distribution. Also the two negative transition states were combined to one transition state (negative/neutral/negative – DSD) with an identical threshold distribution (see Annex, Figure 3.3). As the difference in threshold means in the variable "business expectations" was very small the same procedure was applied.

Given the case that the indifference thresholds are linearly related to some other variable *z,* for example company size or some time-varying parameter, in a way:

$$a = -b = \alpha_0 + \alpha_1 z$$

The probability distribution function that we receive substituting for *a* and *b* in equation (3) with $\alpha_0 + \alpha_1 z$, is:

$$\begin{aligned} \Pr(U \mid x) &= 1 - \Phi\left(\frac{a_0 - Y^* + \alpha_1 z}{\sigma}\right) \\ \Pr(S \mid x) &= \Phi\left(\frac{a_0 - Y^* + \alpha_1 z}{\sigma}\right) - \Phi\left(\frac{a_0 - Y^* + \alpha_1 z}{\sigma}\right) \\ \Pr(D \mid x) &= \Phi\left(\frac{a_0 - Y^* + \alpha_1 z}{\sigma}\right) \end{aligned} \tag{4}$$

The variable *z* can influence the probability distribution in such a way that particular companies in a sample are less likely than other to report rise or fall given the same percentage change of output. Also if the variable *z* is a time-varying parameter, as for example the macroeconomic level of business confidence or macroeconomic uncertainty, the results of the balance statistics will be systematically biased. The study tests whether the indifference thresholds of business expectations depend upon a time-varying factor or significantly vary across respondents by estimating the equation $a=\delta_0+ \delta_1 z_1 \ldots + \delta_n z_n$ $(-b=\delta_0+ \delta_1 z \ldots + \delta_n z_n)$.

4. Estimation Methods and Results

4.1. Testing for Normality of Business Expectations

One of the central assumptions in the modeling of business expectations elicited with three-category scales is the normality and symmetry of the distribution of respondents' expectations about changes in the respective variables. In Chapter 2 of the thesis it has been already demonstrated that both, the distribution of assessments on the present business situation and the distribution of business expectations are skewed in particular periods of the business cycle (see Figures 2.9 and 2.10 in the

Annex to Chapter 2). Also the results of numerical normality tests reveal that the VAS measured business confidence variables are not normally distributed, but the deviation from normality is moderate, particularly in the distribution of the assessments on the present business state (Table 3.3). The skewness and kurtosis of the distribution of the VAS assessments on the present business situation are close to zero and 3 respectively. The deviation from normality is stronger and appears more often in the distribution of the VAS business expectations (Table 3.4). The highest degree of negative skewness of -0.62 has been recorded in May 2006. These results confirm that business expectations are not always symmetrically distributed, but are likely to be skewed in different periods of the business cycle, as suggested by Foster and Gregory (1975). Shapiro-Francia test for normality rejects in all months the null hypothesis that the data are normally distributed for both variables. Similar results deliver the Shapiro-Wilk tests for normality.

These findings have also implications on the following analysis. Most statistical methods request normally distributed data. The normality tests have detected departures from normality. However, in both variables the skewness ranges between zero and -0.6 indicating that the data set is relatively symmetrical. Given the big sample size, the normality tests may detect statistically significant but trivial departures from normality that will have no serious effects on the statistical methods applied in the present study. However, statistical methods not requiring normal distribution are supplementary applied, to assure more powerful tests and valid interpretations.

4.2. Testing the Assumptions about the Indifference Thresholds of Three-category Based Business Expectations

4.2.1. Symmetrical Properties of the Indifference Thresholds

To test the symmetry of the two indifference thresholds, for each of the two variables the mean upper threshold was contrasted with the mean lower threshold at the cross-

sectional level. The difference between the mean upper and the mean lower threshold value was tested against zero with a two-tailed t-test.

As the normal distribution assumption may be violated in some months, another symmetry test is applied that does not require the data to be normally distributed. Wilcoxon matched-pairs rank sum test, also known as the Mann-Whitney-Wilcoxon (MWW) two-sample statistic, tests the hypothesis that two samples are from populations with the same distribution. The MWW rank sum test is a nonparametric test. It may be thought of as performing a two-sample t-test on the data after ranking over the combined sample. Its central assumption is that the populations from which the two samples were taken may differ in their means or medians, but not in their distributional shape. This assumption appears at the first glance realistic (see Figure 3.3 in the Annex). Also the two-sample Kolmogorov-Smirnov test on the equality of distributions reveals that this assumption holds for the distribution of indifference thresholds in all months in the variable "business situation" and sixteen months for the variable "business expectations".

The results of the t-test and the MWW test are summarized in Table 3.5 for the variable "present business situation" and in Table 3.6 for the variable "business expectations". In the assessment of the present business state no significant asymmetry in the upper and the lower threshold was found, except in March 2006. Here, both, the t-test and the MWW test confirm significant results (Table 3.5). In the variable business expectations, in contrast, the thresholds were asymmetrical in most survey waves in 2007 and 2008 (Table 3.6). The results of the t-test and the non-parametric estimates are similar, although the MWW records more often significant asymmetry than the t-test estimates.

The upper threshold (USU) of business expectations is significantly farther away from zero than the lower threshold (DSD), meaning that expected positive changes of business conditions have to exceed a considerably higher threshold before firms report that they expect an improvement than expected negative changes that would make them report an expected deterioration (Figure 3.4).

Such an asymmetry in business expectations shows that respondents may react to the same amplitude of expected gains and losses differently: weighting future losses stronger than future gains. These findings are in line with one of the consequences from the prospect theory of Kahneman und Tversky (1981), that responses to losses are more extreme than the responses to gains.

In the variable "present business situation" there is no systematic asymmetry between the two thresholds (Figure 3.5), meaning that when respondents describe the current state of their business, they do not overstate bad economic situation compared to good. The response pattern in business expectations, in contrast, which involves uncertainty about future business conditions, indicates the presence of loss-aversion.

4.2.2. Variation of the Indifference Thresholds over Time

- *Cross-sectional analysis: Deviation of the thresholds from their long-term mean*

Stability of the indifference thresholds over time is evaluated by several methods. First it was tested whether the mean upper thresholds and the mean lower thresholds in each of the 33 months significantly diverge from their long-term average. This test was performed for both variables – business situation and business expectations. Again t-test and the non-parametric Wilcoxon test were applied to account for non-normality of distributions. Wilcoxon matched-pairs signed-ranks test tests the null hypothesis that the median of the differences between the variable and the expression (the long-term average over 33 months) is zero; no further assumptions are necessary about the distributions.

The results from the analysis of the stability of indifference thresholds over time are summarized in Table 3.7 and Table 3.8 for the variable "present business situation" and in Table 3.9 and Table 3.10 for the variable "business expectations". Results indicate that both the upper and the lower thresholds of the business judgment and the business expectations significantly vary over time. The results of the non-

parametric Wicoxon singned-ranks test are again similar to the results of the t-test, although a significant deviation from the long-time mean is recorded more often with the non-parametric test. The variation of the thresholds over time is much stronger in business expectations (Table 3.9 and Table 3.10) than in the variable present business situation (Table 3.7 and Table 3.8). These results indicate that particularly in the variable "business expectations" the upper and the lower thresholds may depend upon an unknown time-varying parameter.

- *Time-series analysis: Modeling the thresholds as depending on the level of business confidence and a measure of uncertainty*

In past studies the indifference thresholds were modeled as depending on the level of the variable and some measure of uncertainty (Batchelor, 1985, 1986, 1988 and Pesaran, 1984). Uncertainty was as a rule proxied by the dispersion of the variable in the survey at a particular date. Results in Chapter 2 demonstrated that the dispersion of VAS business expectations is positively correlated with the "epistemic" uncertainty in the VAS responses and can be used as proxy for macroeconomic uncertainty. Also the level of business situation and business expectations of the previous month were readily available. The resulted time-series that are used in the following analysis are summarized in Table 3.11.

The upper and the lower thresholds as well as the asymmetry (calculated as balance of the mean upper and the mean lower thresholds) were regressed on the proxy of uncertainty and the aggregate level of the present business situation and the aggregate business expectations. For the variable "present business situation" the regression results are summarized in Table 3.12, for the variable "business expectations" in Table 3.13.

Uncertainty appears to broaden the indifference interval (it has a positive effect on the upper threshold and a negative effect on the lower threshold), particularly in the variable "present business situation". These results indicate that the higher the macroeconomic uncertainty the earlier respondents turn to the neutral category within the three-category scale and the longer they remain within the neutral state.

There is no significant effect of uncertainty on the asymmetry between the thresholds of the variable “present business situation” (Table 3.12).

Also in the variable “business expectations” uncertainty seems to broaden the indifference interval, although rather due to a positive effect on the upper threshold (USU). Uncertainty was also found to be significantly positive related to the asymmetry between the upper and the lower thresholds of business expectations (Table 3.13). The higher the uncertainty, the more the upper threshold exceeds in magnitude the lower threshold, meaning that the expected positive change has to be stronger to make respondents leave the neutral category for the positive category, than the negative change which would make them choosing the negative category. These results indicate that the higher the macroeconomic uncertainty, controlling for the level of business confidence, the more extreme is the response to future losses compared to the response to future gains. In the previous chapter it has been demonstrated that uncertainty generally increases as economic growth falls. As respondents weight losses stronger than gains, particularly in this period of the business cycle, the economic downturn signaled by business expectations may be overstated. At the one hand this may result in a higher lead of the indicator. Indeed, in previous research a tendency was observed that peaks are signaled by business surveys with a larger lead than troughs (Abberger, 2006). At the other hand, when three-category business expectations are quantified or enter models used for point-forecasts of the GDP, they may bias the point-forecasts downward.

The asymmetry between the indifference thresholds of business expectations seems also to be related to the level of the present economic performance, holding uncertainty constant (Table 3.13). The higher the level of the present business situation, the expected changes have to exceed a higher threshold before firms report that they expect an improvement. Also these findings are inline with theoretical considerations and can be explained with the diminishing marginal utility of output growth.

4.2.3. Variation of the Indifference Thresholds across Respondents

- *Panel analysis: Modeling the thresholds as depending on the level of business confidence, a measure of uncertainty and company size*

To test the hypothesis whether the indifference thresholds vary across respondents, in the next step the thresholds are modeled as depending on the company size. The model is estimated at the micro-data level. To test for robustness of the above discussed results with respect to the model specification also the measure of uncertainty and the level of business confidence were included in the model.

The independent variables that entered the model were the proxy for uncertainty in form of the standard deviation of business expectations at each month, the level of aggregate business situation, the level of aggregate business expectations, company size (measured by the number of employees) and the time dummies to control for the variation of the indifference thresholds over time. The proxy of uncertainty and the level of business situation and business expectations vary only over time, but not across respondents. The company size varies both, within and across respondents. The company size was defined by the number of employees. To avoid that outlier (companies with more than 100,000 employees bias the regression results), eleven company size classes[3] were generated.

The model was estimated by the fixed effects (FE) panel regression. Although the FE estimator ignores the between subjects variation, estimating the within subjects variation and is not as efficient as the random effect estimator, it was selected for several reasons: The random effects estimator would be biased a soon as one of the explanatory variables is correlated with the composite error term. The composite error term is unknown. It can include some characteristics of the respondent or the respondent's firm characteristics. So, for example the firm growth may be related to some unobserved firm characteristics or to respondent specific loss aversion. The F-test following the fixed effects panel regression indicates that there are significant

[3] The 11 company size classes were generated according to the number of employees: (1/49 = 1) (50/99 = 2) (100/149 = 3) (150/199 = 4) (200/249 = 5) (250/349 = 6) (350/499 = 7) (500/649 = 8) (650/999 = 9) (1,000/1,999 = 10) (2,000/200,000 = 11).

individual (firm level) effects (Table 3.14 and Table 3.15). Whether the random effects model can be used, was evaluated with the Hausman[4] test that checks whether the random effects estimate is insignificantly different from the fixed effects estimate. The results of the Hausman test indicated that there may be an omitted variable that is correlated with one of the explanatory variables. Consequently the FE estimator was used. An advantage of the FE estimator is, however, that it controls for all time-invariant variables not just those that were in the model.

The results of the FE panel regression of the indifference threshold on company size, a measure of uncertainty and the aggregate level of business situation and business expectations confirm previous findings (Table 3.14 and Table 3.15).

The level of uncertainty increases the upper bound and decreases the lower bound in both variables. The aggregate level of business situation seems to be positively related to the upper threshold (see Table 3.14). The same pattern applies to business expectations, the higher the level of business expectations the expected changes of business conditions have to exceed a considerably higher threshold before firms report that they expect an improvement. For the lower threshold, in contrast, the higher the aggregate level of business expectations, the smaller the negative changes have to be to make firms report that they expect a deterioration.

Although the company size varies weakly within the companies it was found to be significantly related to the width of the indifference threshold in the variable "present business situation" (see Table 3.14). Increasing number of employees appears to narrow the indifference interval of the variable "present business situation", meaning that for larger companies a smaller percentage change in output or profits, results in reporting a fall on the three-category scale, than for small-scale enterprises. These results are inline with theoretical considerations reported in the literature (Ronning, 1990). In the variable "business expectations" company size was not significantly related to the width or the form of the indifference interval.

[4] The Hausman test tests the null hypothesis that the coefficients estimated by the efficient random effects estimator are the same as the ones estimated by the consistent fixed effects estimator. According to the results of the Hausman test the coefficients were jointly significant (Prob>chi² smaller than 0.05) indicating that there may be an omitted variable that is correlated with one of the explanatory variables.

- *Cross-sectional analysis: Modeling the thresholds as depending on the level of business confidence and company size*

To test whether these results hold also at the cross-sectional level the upper and the lower thresholds were modeled as depending on the company size, the firm's business situation appraisal and the firm's business expectations. The uncertainty proxy could not be added to the model as this variable does not vary within months only across months. Company size is expected to have rather a negative sing in the regression of the upper threshold and a positive sing in the regression of the lower threshold.

The results appear to be robust and consistent with the above findings. In the regression of the upper threshold of the variable "business situation" company size has a negative sign in most months and is in two months significant (see Table 3.16). In the regression of the lower threshold of the variable "business situation" company size was found to be three times significantly positive (see Table 3.17). Somewhat contradictory are the results of the regressions of the upper threshold of business expectations on company size and controls (see Table 3.18). Although in the majority of months company size has a negative sign, it is twice significantly positive and only once significantly negative. However, in the regression of the lower threshold of the variable "business expectations" company size was found to be a significant predictor twice, both times with a positive sign (see Table 3.19). The overall results justify the conclusion that for larger companies a smaller expected change may result in choosing the positive or the negative category than for smaller companies.

5. Summary

This chapter was dedicated to the tests of popular assumptions about three-category based economic expectations: symmetry of indifference thresholds, time-constancy and stability across respondents. All three assumptions have been found violated. Furthermore, the normality of the distribution of business expectations was also

found violated in particular periods of the business cycle, although to a degree that may not have serious impacts on statistical inferences.

In particular, the symmetry of the indifference thresholds is not guaranteed in business expectations. The upper threshold of business expectations is significantly farther away from zero than the lower threshold, meaning that expected changes of business conditions have to exceed a considerably higher threshold before firms report that they expect an improvement than expected negative changes that would make them report an expected deterioration. Such an asymmetry in business expectations shows that respondents may react to the same amplitude of expected gains and losses differently: weighting future losses stronger than future gains. These findings are in line with one of the consequences from the prospect theory of Kahneman und Tversky (1981) that includes that responses to losses are more extreme than the responses to gains.

Stability of the indifference thresholds over time was evaluated by several methods. The results indicate that the indifference thresholds also vary over time and depend upon time-varying parameters. As in previous studies, the indifference thresholds were modeled as depending on the level of the variable and some measure of uncertainty. Uncertainty was proxied by the monthly dispersion of business expectations. Uncertainty was found to broaden the indifference interval (having a positive effect on the upper threshold and a negative effect on the lower threshold). These results indicate that the higher the macroeconomic uncertainty the earlier respondents turn to the neutral category within the three-category scale and they remain within the neutral state until the reference variable (output or profits) overcomes a considerably higher threshold than would be necessary in periods of low macroeconomic uncertainty. Uncertainty was also found to be significantly positive related to the asymmetry between the upper and the lower thresholds of business expectations. These results lead to the conclusion that the higher the macroeconomic uncertainty, the more extreme is the response to future losses compared to the response to future gains, meaning that uncertainty make respondents deflate future gains and losses differently. As uncertainty increases as economic

growth falls, these findings may explain the empirical evidence that peaks are signaled by business surveys with a larger lead than troughs.

The asymmetry between the indifference thresholds of business expectations seems also to be related to the level of the present economic performance. The higher the level of the present business situation, the higher the threshold, which expected changes have to exceed before firms report that they expect an improvement. These results may be explained with the diminishing marginal utility of welfare.

The indifference threshold were also found to vary across respondents. Company size appears rather to narrow the indifference interval, meaning that for larger companies a smaller percentage change, results in reporting a fall or a rise on the three-category scale, than for small-scale enterprises. These results are inline with theoretical considerations reported in the literature and hold also at the cross-sectional level, although strong evidence was not found in every month.

The empirical findings demonstrated that popular assumptions made in the analysis of three-category based business expectations do not hold. Consequently, statistical methods that can model these kind of data are to be preferred over methods that rely on these assumptions. For the quantification of business expectations, Pesaran's (1984) regression approach or the refined Carlson-Parkin method is more appropriate than the balance statistics or the original Carlson-Parkin quantification.

On the other hand, the VAS offers a direct measurement instrument of business expectations that is not constricted by strong assumptions. Although it is a qualitative measurement method of business expectations, its measurement scale comes close to an interval measurement and enables an unbiased calculation of the central tendency of the distribution. However, also used simultaneously with the three-category scale, the VAS can help to detect asymmetries and time-variant effects in responses in time. Particularly, the shape and the variation of the indifference interval over time appears to contain cyclical information, which may be explored in more detail in the future.

Annex Chapter 3

Table 3.1. Valid Responses in the Ifo BTS in the German Manufacturing Sector

Month	N	Percent	Cum.
2005/09	1,134	2.5	2.5
2005/10	1,166	2.5	5.0
2005/11	1,183	2.6	7.6
2005/12	1,144	2.5	10.1
2006/01	1,341	2.9	13.0
2006/02	1,326	2.9	15.9
2006/03	1,398	3.1	18.9
2006/04	1,378	3.0	22.0
2006/05	1,372	3.0	24.9
2006/06	1,310	2.9	27.8
2006/07	1,322	2.9	30.7
2006/08	1,249	2.7	33.4
2006/09	1,319	2.9	36.3
2006/10	1,308	2.9	39.1
2006/11	1,336	2.9	42.0
2006/12	1,378	3.0	45.0
2007/01	1,398	3.1	48.1
2007/02	1,347	2.9	51.0
2007/03	1,423	3.1	54.1
2007/04	1,396	3.0	57.2
2007/05	1,404	3.1	60.2
2007/06	1,407	3.1	63.3
2007/07	1,346	2.9	66.2
2007/08	1,299	2.8	69.1
2007/09	1,350	2.9	72.0
2007/10	1,381	3.0	75.0
2007/11	1,356	3.0	78.0
2007/12	1,411	3.1	81.0
2008/01	1,431	3.1	84.2
2008/02	1,465	3.2	87.4
2008/03	1,455	3.2	90.5
2008/04	1,473	3.2	93.7
2008/05	1,407	3.1	96.8
2008/06	1,470	3.2	100.0

Figure 3.1. Density of Responses to the Three-category Scale and the VAS for the Variables "Business Situation" and "Business Expectations"(pooled panel data)

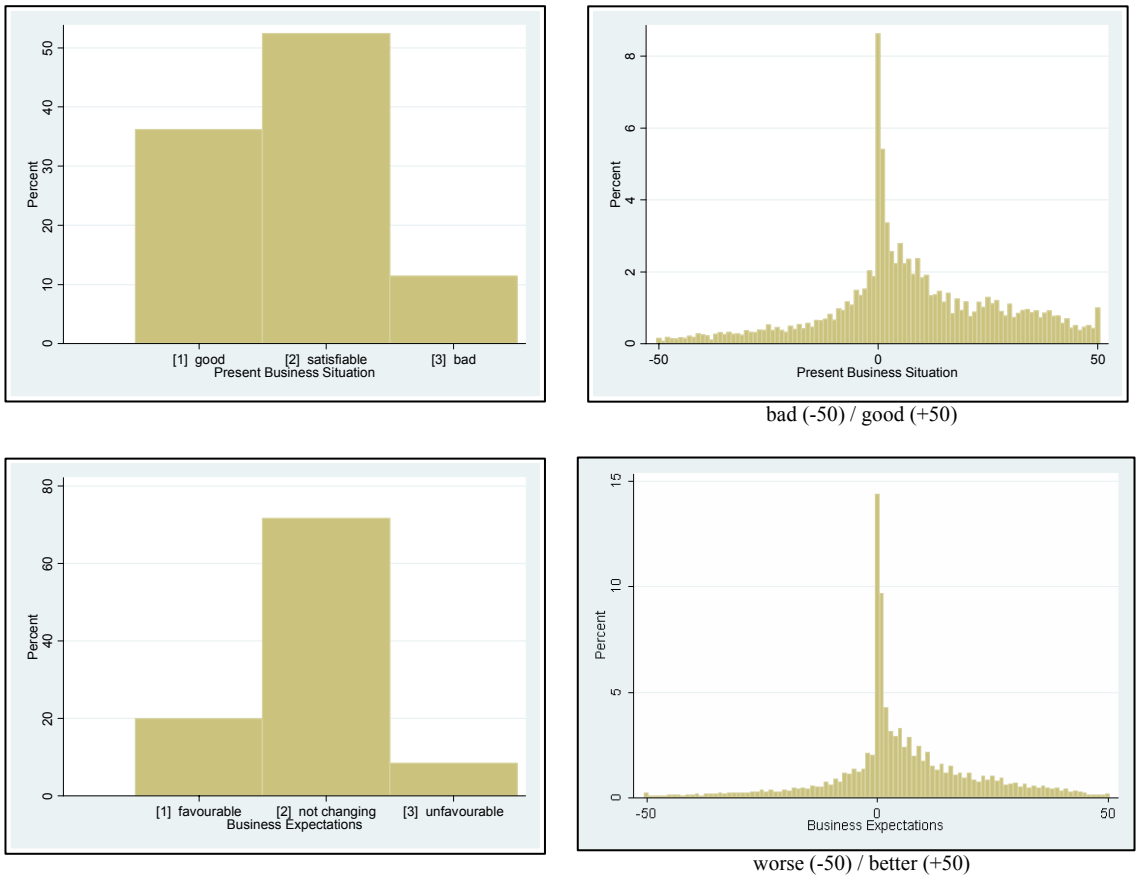


Figure 3.2. Transition States within the Three-category Scale and the Corresponding Distribution of Indifference Thresholds on the VAS (pooled panel data)

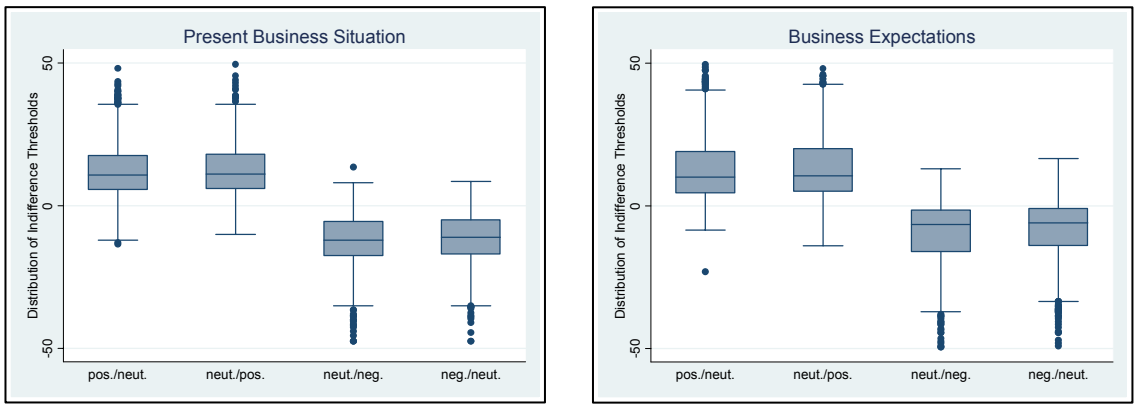


Table 3.2. Results of the t-test on the Equality of Means between the Thresholds within the Transition States

Transition sates	N	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]	Pr(|T| > |t|) =
Business Situation							
pos./neut.	1,810	13.78	0.21	8.95	13.37	14.19	
neut./pos.	1,700	14.08	0.22	8.88	13.66	14.51	
Difference in means		**-0.30**	0.30		-0.89	0.29	***0.32***
neut./neg.	888	-14.32	0.33	9.72	-14.97	-13.68	
neg./neut.	897	-14.14	0.33	9.83	-14.79	-13.50	
Difference in means		**-0.18**	0.46		-1.09	0.73	***0.69***
Business Expectations							
pos./neut.	2,231	15.62	0.23	10.84	15.17	16.07	
neut./pos.	2,354	16.33	0.23	10.94	15.89	16.78	
Difference in means		**-0.71**	0.32		-1.35	-0.08	***0.03***
neut./neg.	1,004	-13.44	0.37	11.80	-14.17	-12.70	
neg./neut.	1,049	-11.36	0.35	11.39	-12.05	-10.67	
Difference in means		**-2.07**	0.51		-3.08	-1.07	***0.00***

Figure 3.3. Distribution of the Negative (blue) and the Positive (orca) Indifference Thresholds on the VAS for the Variables "Business Situation" and "Business Expectations" (pooled panel data)

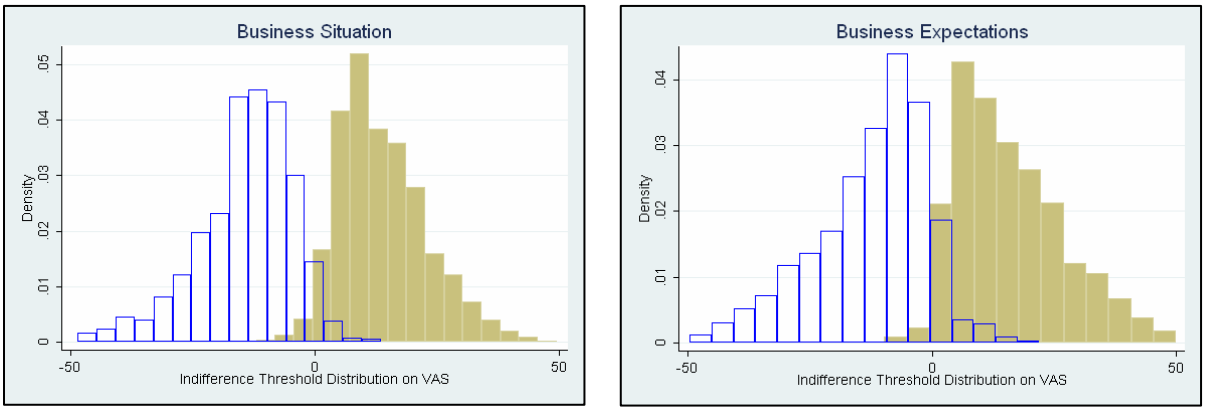


Table 3.3. Summary Statistics of the Variable "Present Business Situation" Measured by the VAS

Month	N	Mean	Med.	SD	Skew.	Kurt.
2005/09	978	-1.30	-2	23.07	0.03	2.30 *
2005/10	995	0.74	2	22.10	-0.04	2.44 *
2005/11	993	-0.54	-2	22.15	0.04	2.56 *
2005/12	985	2.07	3	21.33	-0.07	2.78
2006/01	1,114	2.81	5	21.22	-0.12	2.71
2006/02	1,120	3.68	4	21.08	-0.07	2.77
2006/03	1,174	5.18	5	21.38	-0.16 *	2.72 *
2006/04	1,178	7.69	8	20.80	-0.25 *	2.84
2006/05	1,157	5.69	6	21.47	-0.27 *	2.85
2006/06	1,098	7.59	8	21.38	-0.34 *	2.89
2006/07	1,119	7.93	8	21.63	-0.36 *	2.89
2006/08	1,038	7.62	8	21.57	-0.35 *	2.90
2006/09	1,091	8.87	9	21.36	-0.37 *	2.88
2006/10	1,097	10.22	10	21.06	-0.35 *	2.87
2006/11	1,124	10.39	9	20.84	-0.42 *	3.03
2006/12	1,175	12.03	10	20.81	-0.38 *	3.06
2007/01	1,196	11.24	10	20.36	-0.31 *	2.97
2007/02	1,130	11.41	10	20.11	-0.39 *	3.12
2007/03	1,223	11.61	11	20.12	-0.34 *	3.03
2007/04	1,191	12.45	12	20.71	-0.41 *	2.92
2007/05	1,187	11.47	11	20.90	-0.39 *	2.98
2007/06	1,165	12.13	12	21.13	-0.38 *	2.83
2007/07	1,164	11.92	11	21.18	-0.37 *	2.79
2007/08	1,102	12.03	11	21.08	-0.33 *	2.90
2007/09	1,134	10.92	10	20.44	-0.27 *	2.88
2007/10	1,169	10.25	9	21.43	-0.28 *	2.73 *
2007/11	1,111	10.53	9	21.19	-0.30 *	2.86
2007/12	1,165	9.42	8	21.41	-0.23 *	2.87
2008/01	1,189	9.14	8	21.17	-0.22 *	2.81
2008/02	1,187	10.08	9	20.74	-0.24 *	2.85
2008/03	1,202	9.85	9	20.71	-0.19 *	2.87
2008/04	1,225	8.96	8	21.01	-0.18 *	2.77
2008/05	1,157	8.57	8	21.18	-0.23 *	2.75
2008/06	1,224	7.90	6	21.53	-0.17 *	2.70 *

* Significant deviation from normal distribution $p < 0.05$ level according to the skewness and kurtosis test for normality. **N**-number of observations, **Med.**-median, **SD**-standard deviation, **Skew**.-skewness, **Kurt**.-kurtosis;

Table 3.4. Summary Statistics of the Variable "Business Expectations" Measured by the VAS

Month	N	Mean	Med.	SD	Skew.	Kurt.
2005/09	808	1.62	4	19.32	-0.33 *	3.07
2005/10	821	1.70	4	18.92	-0.22 *	3.13
2005/11	871	0.35	3	18.52	-0.28 *	3.32
2005/12	840	2.28	4	18.21	-0.41 *	3.82 *
2006/01	1,013	5.59	7	17.27	-0.47 *	3.76 *
2006/02	1,010	6.64	7	16.93	-0.45 *	3.93 *
2006/03	1,052	7.14	7	16.98	-0.52 *	3.98 *
2006/04	1,049	8.24	8	16.88	-0.46 *	4.04 *
2006/05	1,034	6.29	7	17.46	-0.62 *	4.01 *
2006/06	971	6.93	7	17.41	-0.48 *	3.96 *
2006/07	1,004	7.04	7	17.60	-0.44 *	3.83 *
2006/08	931	6.71	7	17.68	-0.43 *	3.86 *
2006/09	957	6.05	6	18.15	-0.47 *	3.67 *
2006/10	951	5.63	6	17.19	-0.40 *	3.52 *
2006/11	977	5.87	6	17.76	-0.32 *	3.50 *
2006/12	1,044	7.80	7	16.60	-0.32 *	4.00 *
2007/01	1,042	9.95	9	16.51	-0.34 *	3.80 *
2007/02	1,035	9.94	9	16.46	-0.40 *	3.99 *
2007/03	1,073	10.37	9	16.73	-0.36 *	3.90 *
2007/04	1,043	10.41	9	17.56	-0.35 *	3.70 *
2007/05	1,038	10.10	9	17.13	-0.37 *	3.92 *
2007/06	1,012	9.69	8	17.14	-0.35 *	3.71 *
2007/07	1,006	9.86	8	17.60	-0.33 *	3.70 *
2007/08	954	9.14	8	17.73	-0.39 *	3.89 *
2007/09	978	7.22	7	17.99	-0.30 *	3.59 *
2007/10	1,006	6.10	5	18.20	-0.17 *	3.27 *
2007/11	962	5.30	6	18.01	-0.27 *	3.52 *
2007/12	1,043	6.12	6	18.28	-0.29 *	3.67 *
2008/01	1,048	6.69	6	18.23	-0.30 *	3.45 *
2008/02	1,069	7.43	6	17.75	-0.26 *	3.47 *
2008/03	1,046	7.13	7	17.82	-0.31 *	3.48 *
2008/04	1,063	6.71	7	18.12	-0.28 *	3.36 *
2008/05	1,006	5.41	5	18.57	-0.26 *	3.30
2008/06	1,089	5.14	5	18.54	-0.20 *	3.07

* Significant deviation from normal distribution $p < 0.05$ level according to the skewness and kurtosis test for normality. **N**-number of observations, **Med.**-median, **SD**-standard deviation, **Skew**.-skewness, **Kurt**.-kurtosis;

Table 3.5. Symmetry Test of the Indifference Thresholds for the Variable "Business Situation"

Month	Thresholds in the transition states USU	DSD	Asymmetry (USU+DSD)	Two-tailed t-Test Statistics Ho: Difference=0 t-value	*Sig.*	N	Non-parametric estimates Wilcoxon rank-sum test z-value	*Sig.*
2005/10	15.28	-16.89	**-1.6**	1.16	*0.25*	183	1.29	*0.20*
2005/11	14.66	-15.07	**-0.4**	0.29	*0.77*	142	0.30	*0.77*
2005/12	12.51	-14.62	**-2.1**	1.50	*0.14*	139	1.64	*0.10*
2006/01	13.64	-14.25	**-0.6**	0.41	*0.68*	158	-0.15	*0.88*
2006/02	12.92	-12.92	**0.0**	0.00	*1.00*	176	0.16	*0.87*
2006/03	14.62	-11.68	**2.9**	-2.35	*0.02*	194	-2.43	*0.02*
2006/04	15.12	-14.09	**1.0**	-0.75	*0.45*	216	-1.16	*0.25*
2006/05	13.48	-13.15	**0.3**	-0.24	*0.81*	210	-0.31	*0.76*
2006/06	13.07	-13.25	**-0.2**	0.11	*0.91*	176	-0.56	*0.58*
2006/07	13.76	-15.46	**-1.7**	1.03	*0.31*	178	0.68	*0.50*
2006/08	14.98	-14.51	**0.5**	-0.28	*0.78*	150	-0.48	*0.63*
2006/09	14.70	-13.22	**1.5**	-0.93	*0.36*	158	-1.49	*0.14*
2006/10	15.24	-15.86	**-0.6**	0.34	*0.73*	157	0.13	*0.90*
2006/11	14.68	-15.88	**-1.2**	0.68	*0.50*	149	0.37	*0.71*
2006/12	13.31	-16.56	**-3.3**	1.50	*0.14*	146	1.40	*0.16*
2007/01	13.95	-12.70	**1.3**	-0.70	*0.49*	173	-1.08	*0.28*
2007/02	13.35	-12.07	**1.3**	-0.83	*0.41*	178	-1.26	*0.21*
2007/03	12.35	-12.41	**-0.1**	0.03	*0.98*	151	-0.11	*0.91*
2007/04	14.18	-13.47	**0.7**	-0.47	*0.64*	177	-0.48	*0.63*
2007/05	14.95	-14.89	**0.1**	-0.04	*0.97*	159	0.10	*0.92*
2007/06	14.36	-12.60	**1.8**	-1.19	*0.24*	139	-0.84	*0.40*
2007/07	13.77	-15.19	**-1.4**	0.84	*0.40*	144	1.10	*0.27*
2007/08	14.75	-12.07	**2.7**	-1.68	*0.10*	145	-1.65	*0.10*
2007/09	13.04	-12.08	**1.0**	-0.62	*0.54*	150	-0.20	*0.84*
2007/10	14.75	-13.59	**1.2**	-0.77	*0.44*	151	-0.07	*0.94*
2007/11	14.25	-14.70	**-0.5**	0.27	*0.79*	146	0.25	*0.80*
2007/12	12.35	-13.29	**-0.9**	0.58	*0.56*	143	0.43	*0.67*
2008/01	13.66	-13.89	**-0.2**	0.15	*0.88*	184	-0.41	*0.68*
2008/02	13.42	-16.07	**-2.7**	1.43	*0.16*	167	0.66	*0.51*
2008/03	12.55	-12.42	**0.1**	-0.08	*0.94*	176	-0.78	*0.43*
2008/04	12.71	-14.35	**-1.6**	1.00	*0.32*	168	0.58	*0.56*
2008/05	13.79	-15.62	**-1.8**	1.25	*0.21*	188	1.50	*0.13*
2008/06	13.89	-15.96	**-2.1**	1.31	*0.19*	174	1.21	*0.23*

Each line represents the results of a t-test and the Wilcoxon rank-sum test on the equality of the upper and the lower threshold, testing the asymmetry between the thresholds against zero.

Table 3.6. Symmetry Test of the Indifference Thresholds for the Variable "Business Expectations"

Month	Mean thresholds in the transition states USU	DSD	Asymmetry (USU+DSD)	Two-tailed t-Test Statistics Ho: Difference=0 t-value	*Sig.*	N	Non-parametric estimates Wilcoxon rank-sum test z-value	*Sig.*
2005/10	14.18	-14.32	**-0.1**	0.08	*0.94*	162	0.07	*0.94*
2005/11	12.32	-12.37	**-0.1**	0.03	*0.98*	194	0.24	*0.81*
2005/12	12.34	-12.99	**-0.6**	0.41	*0.68*	184	0.33	*0.74*
2006/01	13.05	-16.83	**-3.8**	2.15	*0.03*	197	1.67	*0.10*
2006/02	12.70	-12.72	**0.0**	0.01	*0.99*	227	0.09	*0.93*
2006/03	13.47	-15.56	**-2.1**	1.39	*0.17*	231	1.23	*0.22*
2006/04	15.01	-12.73	**2.3**	-1.29	*0.20*	211	-1.49	*0.14*
2006/05	15.08	-12.37	**2.7**	-1.63	*0.11*	246	-2.03	*0.04*
2006/06	15.86	-11.95	**3.9**	-2.40	*0.02*	210	-2.35	*0.02*
2006/07	16.61	-10.97	**5.6**	-3.04	*0.00*	210	-3.88	*0.00*
2006/08	15.80	-13.30	**2.5**	-1.47	*0.14*	213	-1.80	*0.07*
2006/09	16.45	-13.87	**2.6**	-1.23	*0.22*	177	-2.05	*0.04*
2006/10	15.69	-12.86	**2.8**	-1.57	*0.12*	192	-1.82	*0.07*
2006/11	13.78	-12.99	**0.8**	-0.47	*0.64*	183	-0.95	*0.34*
2006/12	15.73	-12.37	**3.4**	-1.86	*0.07*	194	-2.35	*0.02*
2007/01	15.43	-12.90	**2.5**	-1.40	*0.16*	222	-1.69	*0.09*
2007/02	15.43	-11.60	**3.8**	-1.84	*0.07*	213	-2.03	*0.04*
2007/03	17.35	-12.76	**4.6**	-2.23	*0.03*	217	-2.63	*0.01*
2007/04	16.89	-11.81	**5.1**	-2.90	*0.00*	223	-2.81	*0.01*
2007/05	16.07	-11.60	**4.5**	-2.46	*0.02*	207	-2.68	*0.01*
2007/06	17.14	-13.26	**3.9**	-2.01	*0.05*	195	-2.39	*0.02*
2007/07	18.28	-11.32	**7.0**	-3.64	*0.00*	203	-3.58	*0.00*
2007/08	17.59	-12.18	**5.4**	-2.58	*0.01*	183	-2.91	*0.00*
2007/09	16.58	-12.33	**4.2**	-2.26	*0.03*	205	-2.45	*0.01*
2007/10	17.41	-13.36	**4.1**	-2.04	*0.04*	182	-2.51	*0.01*
2007/11	17.61	-9.86	**7.8**	-4.52	*0.00*	196	-4.37	*0.00*
2007/12	17.80	-11.29	**6.5**	-3.58	*0.00*	178	-4.01	*0.00*
2008/01	18.10	-9.71	**8.4**	-5.56	*0.00*	223	-5.39	*0.00*
2008/02	16.72	-9.51	**7.2**	-4.74	*0.00*	211	-4.92	*0.00*
2008/03	17.10	-9.20	**7.9**	-5.44	*0.00*	208	-5.21	*0.00*
2008/04	15.97	-11.04	**4.9**	-3.00	*0.00*	229	-2.61	*0.01*
2008/05	16.40	-11.94	**4.5**	-2.45	*0.02*	201	-2.55	*0.01*
2008/06	17.86	-12.38	**5.5**	-3.76	*0.00*	239	-3.60	*0.00*

Each line presents the results of a t-test and the Wilcoxon rank-sum test on the equality of the upper and the lower threshold, testing the asymmetry between the thresholds against zero.

Figure 3.4. The Upper and the Lower Thresholds in the Variable "Business Expectations" and the Asymmetry between the Thresholds (yellow bars) in 33 Consecutive Survey Months

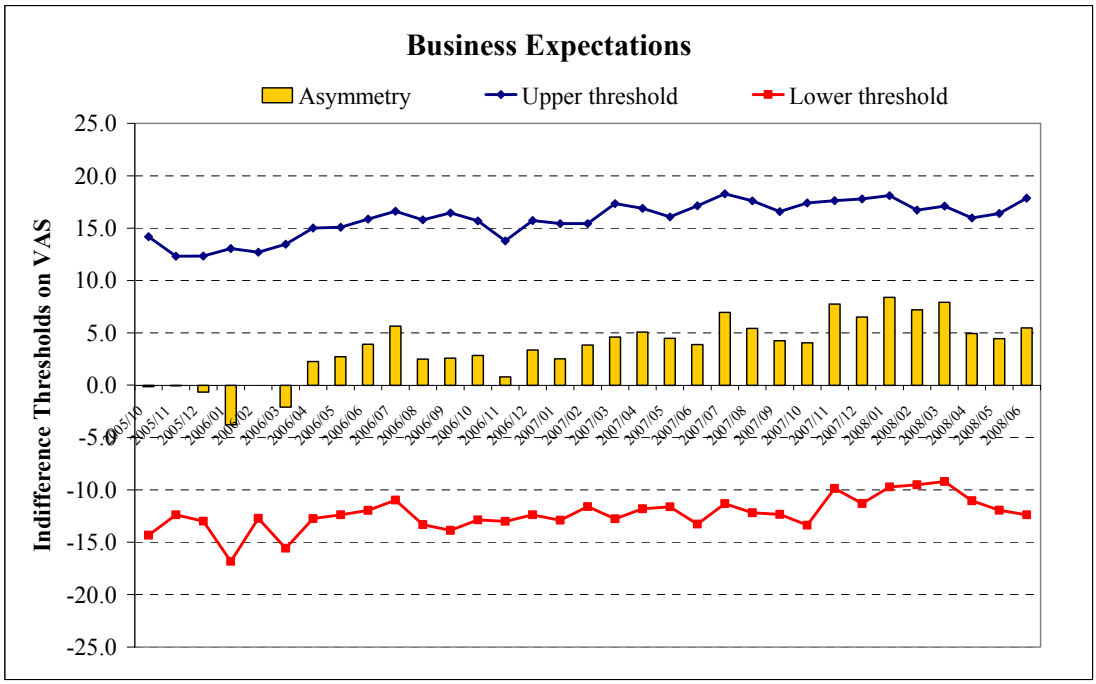


Figure 3.5. The Upper and the Lower Thresholds in the Variable "Present Business Situation" and the Asymmetry between the Thresholds (yellow bars) in 33 Consecutive Survey Months

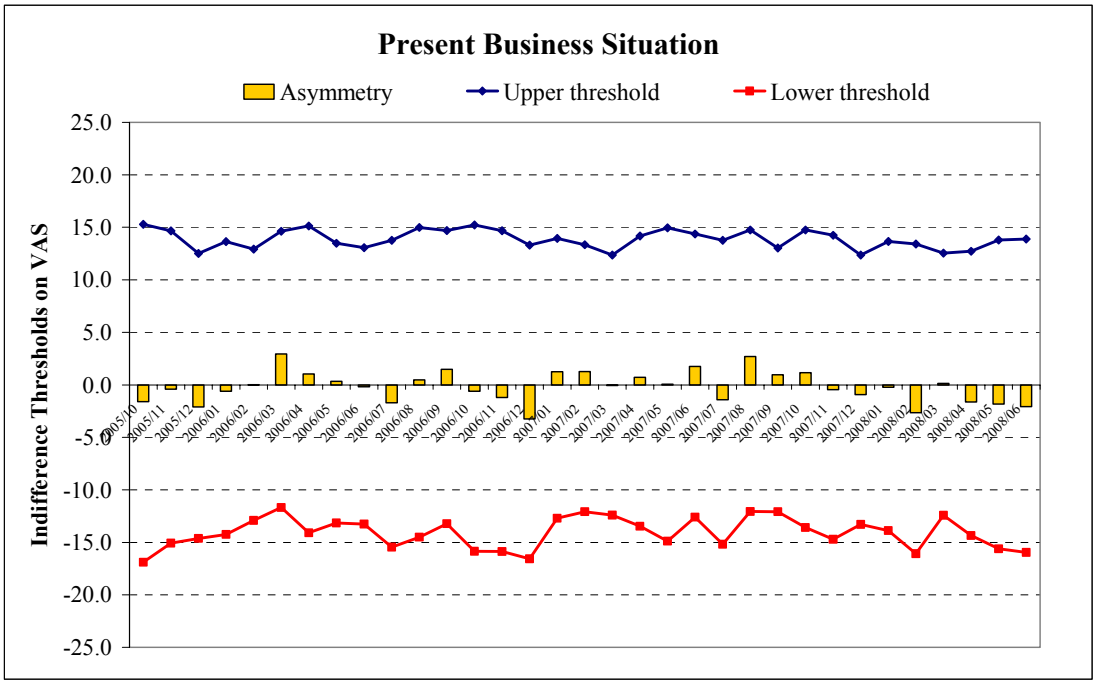


Table 3.7. Deviation of the Upper Threshold (USU) from its Long-time Mean in the Variable "Business Situation"

	Deviation of the threshold from its long-time mean: 13.86		Two-tailed t-test statistics Ho: Difference=0		Non-parametric estimates: Wilcoxon signed-ranks test	
Month	**USU**	**Divergence**	**t-value**	***Sig.***	***Sig.***	**N**
2005/10	15.28	1.42	1.35	*0.18*	*0.82*	183
2005/11	14.66	0.80	0.77	*0.44*	*0.63*	142
2005/12	12.51	-1.35	-1.31	*0.20*	*0.02*	139
2006/01	13.64	-0.23	-0.27	*0.79*	*0.35*	158
2006/02	12.92	-0.94	-1.10	*0.27*	*0.09*	176
2006/03	14.62	0.76	0.92	*0.36*	*1.00*	194
2006/04	15.12	1.26	1.56	*0.12*	*0.25*	216
2006/05	13.48	-0.38	-0.54	*0.59*	*0.01*	210
2006/06	13.07	-0.79	-1.09	*0.28*	*0.06*	176
2006/07	13.76	-0.10	-0.12	*0.90*	*0.53*	178
2006/08	14.98	1.12	1.29	*0.20*	*1.00*	150
2006/09	14.70	0.84	1.01	*0.32*	*0.69*	158
2006/10	15.24	1.38	1.53	*0.13*	*1.00*	157
2006/11	14.68	0.81	0.98	*0.33*	*0.85*	149
2006/12	13.31	-0.56	-0.67	*0.51*	*0.13*	146
2007/01	13.95	0.09	0.13	*0.90*	*0.93*	173
2007/02	13.35	-0.52	-0.69	*0.49*	*0.14*	178
2007/03	12.35	-1.51	-1.82	*0.07*	*0.01*	151
2007/04	14.18	0.32	0.39	*0.69*	*0.86*	177
2007/05	14.95	1.09	1.18	*0.24*	*1.00*	159
2007/06	14.36	0.50	0.54	*0.59*	*0.42*	139
2007/07	13.77	-0.09	-0.09	*0.92*	*0.69*	144
2007/08	14.75	0.89	0.99	*0.32*	*0.31*	145
2007/09	13.04	-0.82	-0.85	*0.40*	*0.05*	150
2007/10	14.75	0.89	0.87	*0.39*	*0.55*	151
2007/11	14.25	0.39	0.41	*0.68*	*0.20*	146
2007/12	12.35	-1.51	-1.87	*0.06*	*0.00*	143
2008/01	13.66	-0.20	-0.27	*0.79*	*0.24*	184
2008/02	13.42	-0.44	-0.63	*0.53*	*0.32*	167
2008/03	12.55	-1.31	-1.59	*0.12*	*0.00*	176
2008/04	12.71	-1.15	-1.56	*0.12*	*0.03*	168
2008/05	13.79	-0.07	-0.07	*0.94*	*0.10*	188
2008/06	13.89	0.03	0.03	*0.98*	*0.10*	174

Each line presents the results of a t-test and the Wilcoxon matched-pairs signed-ranks test testing the null hypothesis that the mean/median of the differences between the variable and the expression (long-time mean) is zero.

Table 3.8. Deviation of the Lower Threshold (DSD) from its Long-time Mean in the Variable "Business Situation"

	Deviation of the threshold from its long-time mean: 14.12		Two-tailed t-test statistics Ho: Difference=0		Non-parametric estimates: Wilcoxon signed-ranks test	
Month	**DSD**	**Divergence**	**t-value**	***Sig.***	***Sig.***	**N**
2005/10	-16.89	-2.76	3.11	*0.00*	*0.01*	183
2005/11	-15.07	-0.94	0.96	*0.34*	*0.35*	142
2005/12	-14.62	-0.49	0.52	*0.61*	*1.00*	139
2006/01	-14.25	-0.13	0.10	*0.92*	*0.54*	158
2006/02	-12.92	1.21	-1.23	*0.22*	*0.30*	176
2006/03	-11.68	2.44	-2.61	*0.01*	*0.00*	194
2006/04	-14.09	0.04	-0.03	*0.97*	*0.33*	216
2006/05	-13.15	0.98	-0.82	*0.41*	*0.07*	210
2006/06	-13.25	0.87	-0.59	*0.56*	*0.13*	176
2006/07	-15.46	-1.34	0.92	*0.36*	*0.68*	178
2006/08	-14.51	-0.39	0.26	*0.79*	*0.07*	150
2006/09	-13.22	0.91	-0.66	*0.51*	*0.10*	158
2006/10	-15.86	-1.74	1.09	*0.28*	*0.89*	157
2006/11	-15.88	-1.75	1.13	*0.26*	*0.88*	149
2006/12	-16.56	-2.44	1.21	*0.23*	*0.49*	146
2007/01	-12.70	1.42	-0.87	*0.39*	*0.56*	173
2007/02	-12.07	2.05	-1.53	*0.13*	*0.12*	178
2007/03	-12.41	1.72	-1.08	*0.29*	*0.10*	151
2007/04	-13.47	0.65	-0.51	*0.61*	*0.34*	177
2007/05	-14.89	-0.76	0.54	*0.59*	*0.77*	159
2007/06	-12.60	1.52	-1.32	*0.19*	*0.43*	139
2007/07	-15.19	-1.06	0.78	*0.44*	*0.36*	144
2007/08	-12.07	2.06	-1.56	*0.13*	*0.03*	145
2007/09	-12.08	2.05	-1.67	*0.10*	*0.01*	150
2007/10	-13.59	0.53	-0.49	*0.63*	*0.89*	151
2007/11	-14.70	-0.58	0.41	*0.69*	*0.88*	146
2007/12	-13.29	0.84	-0.61	*0.55*	*0.09*	143
2008/01	-13.89	0.24	-0.18	*0.86*	*0.10*	184
2008/02	-16.07	-1.95	1.13	*0.26*	*0.24*	167
2008/03	-12.42	1.70	-1.19	*0.24*	*0.06*	176
2008/04	-14.35	-0.23	0.16	*0.88*	*0.40*	168
2008/05	-15.62	-1.49	1.35	*0.18*	*0.81*	188
2008/06	-15.96	-1.84	1.44	*0.15*	*0.40*	174

Each line presents the results of a t-test and the Wilcoxon matched-pairs signed-ranks test testing the null hypothesis that the mean/median of the differences between the variable and the expression (long-time mean) is zero.

Table 3.9. Deviation of the Upper Threshold (USU) from its Long-time Mean in the Variable "Business Expectations"

	Deviation of the threshold from its long-time mean: 15.90		Two-tailed t-test statistics Ho: Difference=0		Non-parametric estimates: Wilcoxon signed-ranks test	
Month	**USU**	**Divergence**	**t-value**	***Sig.***	***Sig.***	**N**
2005/10	14.18	-1.72	-1.61	*0.11*	*0.02*	162
2005/11	12.32	-3.57	-3.31	*0.00*	*0.00*	194
2005/12	12.34	-3.55	-3.97	*0.00*	*0.00*	184
2006/01	13.05	-2.85	-3.47	*0.00*	*0.00*	197
2006/02	12.70	-3.20	-4.42	*0.00*	*0.00*	227
2006/03	13.47	-2.42	-3.68	*0.00*	*0.00*	231
2006/04	15.01	-0.89	-1.26	*0.21*	*0.07*	211
2006/05	15.08	-0.81	-1.10	*0.27*	*0.10*	246
2006/06	15.86	-0.04	-0.04	*0.97*	*0.57*	210
2006/07	16.61	0.71	0.80	*0.42*	*0.11*	210
2006/08	15.80	-0.09	-0.11	*0.92*	*0.05*	213
2006/09	16.45	0.55	0.54	*0.59*	*0.32*	177
2006/10	15.69	-0.20	-0.21	*0.84*	*0.42*	192
2006/11	13.78	-2.12	-2.30	*0.02*	*0.02*	183
2006/12	15.73	-0.17	-0.18	*0.86*	*0.10*	194
2007/01	15.43	-0.47	-0.58	*0.57*	*0.05*	222
2007/02	15.43	-0.46	-0.58	*0.57*	*0.13*	213
2007/03	17.35	1.45	1.69	*0.09*	*0.70*	217
2007/04	16.89	0.99	1.14	*0.26*	*0.70*	223
2007/05	16.07	0.17	0.20	*0.84*	*0.47*	207
2007/06	17.14	1.24	1.33	*0.19*	*0.74*	195
2007/07	18.28	2.38	2.53	*0.01*	*0.26*	203
2007/08	17.59	1.70	1.78	*0.08*	*0.50*	183
2007/09	16.58	0.68	0.76	*0.45*	*0.22*	205
2007/10	17.41	1.52	1.46	*0.15*	*0.42*	182
2007/11	17.61	1.71	1.51	*0.13*	*0.58*	196
2007/12	17.80	1.90	1.63	*0.11*	*0.56*	178
2008/01	18.10	2.21	2.25	*0.03*	*0.36*	223
2008/02	16.72	0.82	0.85	*0.40*	*0.67*	211
2008/03	17.10	1.20	1.33	*0.19*	*0.17*	208
2008/04	15.97	0.07	0.08	*0.93*	*0.04*	229
2008/05	16.40	0.51	0.49	*0.63*	*0.37*	201
2008/06	17.86	1.96	1.90	*0.06*	*0.93*	239

Each line presents the results of a t-test and the Wilcoxon matched-pairs signed-ranks test testing the null hypothesis that the mean/median of the differences between the variable and the expression (long-time mean) is zero.

Table 3.10. Deviation of the Lower Threshold (DSD) from its Long-time Mean in the Variable "Business Expectations"

	Deviation of the threshold from its long-time mean: 12.26		Two-tailed t-test statistics		Non-parametric estimates: Wilcoxon signed-ranks test	
Month	**DSD**	**Divergence**	**t-value**	***Sig.***	***Sig.***	**N**
2005/10	-14.32	-2.06	1.50	*0.14*	*1.00*	162
2005/11	-12.37	-0.11	0.09	*0.93*	*0.34*	194
2005/12	-12.99	-0.73	0.57	*0.57*	*0.51*	184
2006/01	-16.83	-4.57	2.93	*0.00*	*0.39*	197
2006/02	-12.72	-0.46	0.31	*0.76*	*0.58*	227
2006/03	-15.56	-3.31	2.44	*0.02*	*0.90*	231
2006/04	-12.73	-0.47	0.30	*0.77*	*0.77*	211
2006/05	-12.37	-0.11	0.07	*0.94*	*0.21*	246
2006/06	-11.95	0.31	-0.22	*0.82*	*0.37*	210
2006/07	-10.97	1.29	-0.79	*0.43*	*0.01*	210
2006/08	-13.30	-1.05	0.72	*0.47*	*0.26*	213
2006/09	-13.87	-1.61	0.89	*0.38*	*0.42*	177
2006/10	-12.86	-0.60	0.40	*0.69*	*0.54*	192
2006/11	-12.99	-0.73	0.52	*0.61*	*0.21*	183
2006/12	-12.37	-0.11	0.07	*0.94*	*0.16*	194
2007/01	-12.90	-0.64	0.40	*0.69*	*0.14*	222
2007/02	-11.60	0.66	-0.35	*0.73*	*0.04*	213
2007/03	-12.76	-0.50	0.27	*0.79*	*0.38*	217
2007/04	-11.81	0.45	-0.30	*0.77*	*0.03*	223
2007/05	-11.60	0.66	-0.41	*0.68*	*0.15*	207
2007/06	-13.26	-1.00	0.59	*0.56*	*0.32*	195
2007/07	-11.32	0.94	-0.56	*0.58*	*0.11*	203
2007/08	-12.18	0.08	-0.04	*0.96*	*0.08*	183
2007/09	-12.33	-0.07	0.04	*0.96*	*0.79*	205
2007/10	-13.36	-1.10	0.65	*0.52*	*0.35*	182
2007/11	-9.86	2.40	-1.86	*0.07*	*0.02*	196
2007/12	-11.29	0.97	-0.70	*0.49*	*0.06*	178
2008/01	-9.71	2.54	-2.22	*0.03*	*0.00*	223
2008/02	-9.51	2.75	-2.34	*0.02*	*0.00*	211
2008/03	-9.20	3.06	-2.70	*0.01*	*0.00*	208
2008/04	-11.04	1.22	-0.89	*0.38*	*0.07*	229
2008/05	-11.94	0.32	-0.21	*0.83*	*0.14*	201
2008/06	-12.38	-0.12	0.11	*0.91*	*0.15*	239

Each line presents the results of a t-test and the Wilcoxon matched-pairs signed-ranks test testing the null hypothesis that the mean/median of the differences between the variable and the expression (long-time mean) is zero.

Table 3.11. Time-Series of Indifference Thresholds, their Asymmetry and a Set of Macroeconomic Variables (Data Set for the Regressions in Table 3.12 and Table 3.13)

	Present Business Situation			Business Expectations					
Month	**USU**	**DSD**	**Asymmetry**	**USU**	**DSD**	**Asymmetry**	**A**	**B**	**C**
2005/10	15.28	-16.89	-1.60	14.18	-14.32	-0.14	19.32	-1.30	1.62
2005/11	14.66	-15.07	-0.41	12.32	-12.37	-0.05	18.94	0.62	1.55
2005/12	12.51	-14.62	-2.11	12.34	-12.99	-0.64	18.51	-0.51	0.31
2006/01	13.64	-14.25	-0.61	13.05	-16.83	-3.78	18.29	2.01	2.22
2006/02	12.92	-12.92	0.00	12.70	-12.72	-0.01	17.24	2.81	5.70
2006/03	14.62	-11.68	2.94	13.47	-15.56	-2.09	16.96	3.43	6.57
2006/04	15.12	-14.09	1.04	15.01	-12.73	2.27	17.07	5.10	7.05
2006/05	13.48	-13.15	0.34	15.08	-12.37	2.71	16.89	7.61	8.19
2006/06	13.07	-13.25	-0.18	15.86	-11.95	3.91	17.43	5.72	6.28
2006/07	13.76	-15.46	-1.70	16.61	-10.97	5.64	17.38	7.55	6.86
2006/08	14.98	-14.51	0.47	15.80	-13.30	2.50	17.60	7.78	7.00
2006/09	14.70	-13.22	1.48	16.45	-13.87	2.57	17.68	7.40	6.55
2006/10	15.24	-15.86	-0.62	15.69	-12.86	2.83	18.20	8.90	5.87
2006/11	14.68	-15.88	-1.20	13.78	-12.99	0.79	17.15	10.12	5.61
2006/12	13.31	-16.56	-3.26	15.73	-12.37	3.36	17.71	10.29	5.84
2007/01	13.95	-12.70	1.25	15.43	-12.90	2.53	16.66	11.95	7.68
2007/02	13.35	-12.07	1.27	15.43	-11.60	3.84	16.48	11.18	9.95
2007/03	12.35	-12.41	-0.05	17.35	-12.76	4.59	16.42	11.35	9.94
2007/04	14.18	-13.47	0.71	16.89	-11.81	5.08	16.82	11.46	10.27
2007/05	14.95	-14.89	0.06	16.07	-11.60	4.47	17.70	12.20	10.25
2007/06	14.36	-12.60	1.76	17.14	-13.26	3.88	17.15	11.25	10.02
2007/07	13.77	-15.19	-1.42	18.28	-11.32	6.95	17.15	12.07	9.59
2007/08	14.75	-12.07	2.69	17.59	-12.18	5.42	17.68	11.77	9.72
2007/09	13.04	-12.08	0.96	16.58	-12.33	4.24	17.71	11.89	9.09
2007/10	14.75	-13.59	1.16	17.41	-13.36	4.06	18.00	10.86	7.16
2007/11	14.25	-14.70	-0.45	17.61	-9.86	7.75	18.17	10.21	6.05
2007/12	12.35	-13.29	-0.93	17.80	-11.29	6.51	17.98	10.47	5.27
2008/01	13.66	-13.89	-0.23	18.10	-9.71	8.39	18.25	9.41	6.03
2008/02	13.42	-16.07	-2.66	16.72	-9.51	7.20	18.22	9.10	6.71
2008/03	12.55	-12.42	0.13	17.10	-9.20	7.91	17.67	9.92	7.38
2008/04	12.71	-14.35	-1.64	15.97	-11.04	4.94	17.87	9.77	7.15
2008/05	13.79	-15.62	-1.82	16.40	-11.94	4.46	18.10	8.84	6.70
2008/06	13.89	-15.96	-2.08	17.86	-12.38	5.48	18.53	8.54	5.43

A: Uncertainty: Monthly dispersion of VAS business expectations.
B: Aggregate business situation (VAS): Monthly mean of the VAS business situation.
C: Aggregate business expectations (VAS): Monthly mean of the VAS business expectations.

Table 3.12. Results of the Regression of the Time-series of Business Situation Indifference Thresholds and their Asymmetry on a Set of Independent Macroeconomic Variables

	Business Situation		
Set of explanatory variables	**USU**	**DSD**	**Asymmetry**
Uncertainty [a)]	0.61 *	-0.93 *	-0.31
	(0.35)	*(0.48)*	*(0.49)*
Aggregate business situation (VAS)	-0.11	-0.13	-0.24 **
	(0.08)	*(0.10)*	*(0.11)*
Aggregate business expectations (VAS)	0.26 *	0.25	0.51 **
	(0.15)	*(0.20)*	*(0.21)*
Constant	2.20	1.64	3.83
	(6.59)	*(9.09)*	*(9.35)*
F (3, 29)	1.35	6.07	5.66
R²	0.12	0.39	0.37

Each column contains the results of a separate regression of the indifference thresholds (USU and DSD) and their asymmetry (USU+DSD) on a set of explanatory variables. Standard errors in parenthesis. *** significance level <0.01, ** significance level <0.05, * significance level <0.10. The data set is reported in Table 3.11.). a**)** Uncertainty is proxied by the standard deviation of business expectations measured by the VAS in the respective survey month.

Table 3.13. Results of the Regression of the Time-series of Business Expectations Indifference Thresholds and their Asymmetry on a Set of Independent Macroeconomic Variables

	Business Expectations		
Set of explanatory variables	**USU**	**DSD**	**Asymmetry**
Uncertainty [a)]	1.27 ***	0.77	2.04 ***
	(0.38)	*(0.55)*	*(0.74)*
Aggregate business situation (VAS)	0.29 ***	0.24 *	0.53 ***
	(0.08)	*(0.12)*	*(0.16)*
Aggregate business expectations (VAS)	0.30 *	0.04	0.34
	(0.16)	*(0.23)*	*(0.31)*
Constant	-11.0	-28.1 **	-39.1 ***
	(7.20)	*(10.5)*	*(14.0)*
F (3, 29)	24.85	4.11	16.25
R²	0.72	0.30	0.63

Each column contains the results of a separate regression of the indifference thresholds (USU and DSD) and their asymmetry (USU+DSD) on a set of explanatory variables. Standard errors in parenthesis. *** significance level <0.01, ** significance level <0.05, * significance level <0.10. The data set is reported in Table 3.11. a**)** Uncertainty is proxied as the standard deviation of business expectations measured by the VAS in the respective survey month.

Table 3.14. Fixed Effects Panel Regression of the Business Situation Indifference Thresholds on the Aggregate Level of the Business Situation and Business Expectations Measured by the VAS and a Proxy for Uncertainty (Individual and Time Fixed effects)

Explanatory variables	Business Situation	
	Upper Threshold USU	Lower Threshold DSD
Company size class [a)]	-0.16	1.59 ***
	(0.30)	*(0.63)*
Uncertainty [b)]	0.88 *	-1.76 ***
	(0.48)	*(0.68)*
Aggregate business situation (VAS)	0.23 **	-0.04
	(0.11)	*(0.18)*
Aggregate business expectations (VAS)	-0.09	0.01
	(0.20)	*(0.33)*
Constant [c)]	-0.88	10.76
	(9.12)	*(13.00)*
Sigma u	8.24	9.79
Sigma e	5.66	5.58
Rho	0.68	0.76
F test that all u_i=0	5.97 ***	6.46 ***
F(df)	F(1,056; 2,491)	F(674; 1,156)
N	3,581	1,864

Standard errors in parenthesis. *** significance level <0.01, ** significance level <0.05, * significance level <0.10. **a)** Company size is measured by 11 company size classes according to the number of employees: (1/49 = 1) (50/99 = 2) (100/149 = 3) (150/199 = 4) (200/249 = 5) (250/349 = 6) (350/499 = 7) (500/649 = 8) (650/999 = 9) (1000/1999 = 10) (2000/200000 = 11). b**)** Uncertainty is proxied by the standard deviation of business expectations measured by the VAS in the respective survey month. **c)** Additional controls: time dummies.

Table 3.15. Fixed Effects Panel Regression of the Business Expectations Indifference Thresholds on the Aggregate Level of Business Situation and Business Expectations Measured by the VAS and a Proxy for Uncertainty (Individual and Time Fixed effects)

Explanatory variables	Business Expectations	
	Upper Threshold USU	Lower Threshold DSD
Company size class [a)]	0.15	0.61
	(0.27)	*(0.69)*
Uncertainty [b)]	0.98 **	-0.21
	(0.44)	*(1.07)*
Aggregate business situation (VAS)	0.13	-0.05
	(0.10)	*(0.18)*
Aggregate business expectations (VAS)	0.35 *	0.68 *
	(0.19)	*(0.39)*
Constant [c)]	-6.15	-14.94
	(8.30)	*(20.68)*
Sigma u	9.4	10.27
Sigma e	6.24	7.00
Rho	0.69	0.68
F test that all u_i=0	8.93 ***	5.80 ***
	F(1,166; 3,469)	F(722; 1,341)
N	4,669	2,097

Standard errors in parenthesis. *** significance level <0.01, ** significance level <0.05, * significance level <0.10. **a)** Company size is measured by 11 company size classes according to the number of employees: (1/49 = 1) (50/99 = 2) (100/149 = 3) (150/199 = 4) (200/249 = 5) (250/349 = 6) (350/499 = 7) (500/649 = 8) (650/999 = 9) (1000/1999 = 10) (2000/200000 = 11). b**)** Uncertainty is proxied as the standard deviation of business expectations measured by the VAS in the respective survey month. **c)** Additional controls: time dummies.

Table 3.16. Regression of the Business Situation Upper Indifference Threshold (USU) on Company Size and the Individual Level of Business Situation and Business Expectations Measured by the VAS

Month	Company size[a)]	Business situation	Business expect.	Constant	F	R^2
2005/10	0.21	0.46***	0.19**	5.35**	30.21 (3 , 61)	0.60
2005/11	0.06	0.51***	0.09	5.98**	18.26 (3 , 53)	0.51
2005/12	0.21	0.56***	0.05	5.62***	40.56 (3 , 52)	0.70
2006/01	0.14	0.54***	0.13*	4.28**	38.39 (3 , 73)	0.61
2006/02	-0.22	0.50***	0.15*	7.03***	25.2 (3 , 75)	0.50
2006/03	0.20	0.57***	0.02	5.41***	42.15 (3 , 93)	0.58
2006/04	-0.06	0.50***	0.21***	5.38***	77.4 (3 , 111)	0.68
2006/05	-0.18	0.65***	0.07	4.22***	107.92 (3 , 132)	0.71
2006/06	-0.36*	0.50***	0.19**	6.51***	43.93 (3 , 97)	0.58
2006/07	0.04	0.63***	0.00	5.35***	52.76 (3 , 101)	0.61
2006/08	-0.21	0.47***	0.20***	5.64***	68.45 (3 , 86)	0.70
2006/09	-0.3	0.48***	0.11	8.21***	34.12 (3 , 85)	0.55
2006/10	-0.32	0.68***	0.01	6.98***	71.18 (3 , 84)	0.72
2006/11	0.21	0.65***	0.06	3.93**	61.98 (3 , 88)	0.68
2006/12	-0.19	0.65***	0.09*	5.11***	84.22 (3 , 99)	0.72
2007/01	0.30*	0.54***	0.12**	2.97**	87.05 (3 , 111)	0.70
2007/02	-0.15	0.45***	0.23***	5.01***	71.65 (3 , 94)	0.70
2007/03	-0.11	0.58***	0.15**	4.26***	72.49 (3 , 89)	0.71
2007/04	0.07	0.66***	0.13*	4.17***	68.04 (3 , 104)	0.66
2007/05	0.31	0.73***	0.13*	0.09	97.56 (3 , 95)	0.75
2007/06	-0.21	0.62***	0.10	4.85**	56 (3 , 80)	0.68
2007/07	0.27	0.59***	0.18**	2.14	86.43 (3 , 83)	0.76
2007/08	0.34	0.55***	-0.01	3.86**	40.42 (3 , 83)	0.59
2007/09	-0.63***	0.85***	0.00	5.14***	92.57 (3 , 78)	0.78
2007/10	-0.10	0.75***	0.07	2.51	110.5 (3 , 83)	0.80
2007/11	-0.21	0.74***	0.05	4.72***	137.25 (3 , 82)	0.83
2007/12	-0.16	0.62***	0.08	3.50**	52.29 (3 , 78)	0.67
2008/01	0.19	0.51***	0.14***	3.43***	88.18 (3 , 94)	0.74
2008/02	-0.05	0.55***	0.04	5.91***	46.66 (3 , 99)	0.59
2008/03	0.06	0.56***	0.05	5.39***	62.55 (3 , 103)	0.65
2008/04	-0.04	0.57***	-0.01	5.62***	38.99 (3 , 90)	0.57
2008/05	0.20	0.55***	0.18**	3.11*	56.28 (3 , 94)	0.64
2008/06	-0.17	0.62***	0.14*	3.83*	52.55 (3 , 87)	0.64

Each line represents the results of a separate regression of the indifference threshold on the set of explanatory variables reported in columns. **a)** Company size is measured by 11 company size classes according to the number of employees: (1/49 = 1) (50/99 = 2) (100/149 = 3) (150/199 = 4) (200/249 = 5) (250/349 = 6) (350/499 = 7) (500/649 = 8) (650/999 = 9) (1000/1999 = 10) (2000/200000 = 11).

Table 3.17. Regression of the Business Situation Lower Indifference Threshold (DSD) on Company Size and the Individual Level of Business Situation and Business Expectations Measured by the VAS

Month	Company size[a)]	Business situation	Business expect.	Constant	F	R^2
2005/10	0.48*	0.44***	0.11*	-9.93***	35.7 (3 , 78)	0.58
2005/11	0.31	0.39***	0.27***	-8.49***	37.17 (3 , 52)	0.68
2005/12	-0.31	0.49***	0.13**	-3.96*	25.54 (3 , 53)	0.59
2006/01	-0.22	0.54***	0.17**	-4.71**	36.58 (3 , 48)	0.70
2006/02	-0.18	0.49***	0.06	-5.54**	22.09 (3 , 59)	0.53
2006/03	0.30	0.46***	0.03	-7.06***	27.16 (3 , 64)	0.56
2006/04	0.02	0.66***	0.08	-2.47	44.46 (3 , 64)	0.68
2006/05	0.39	0.51***	0.16*	-8.51***	23.88 (3 , 45)	0.61
2006/06	0.09	0.61***	0.13*	-4.88*	34.06 (3 , 36)	0.74
2006/07	0.50	0.51**	0.11	-9.39**	14.34 (3 , 35)	0.55
2006/08	-0.03	0.49***	0.12	-6.36**	23.22 (3 , 35)	0.67
2006/09	-0.1	0.50***	0.22**	-4.28*	46.32 (3 , 43)	0.76
2006/10	0.30	0.59***	0.16	-6.23*	22.82 (3 , 41)	0.63
2006/11	-0.18	0.70***	0.02	-1.91	23.17 (3 , 35)	0.67
2006/12	0.65	0.56***	0.14	-7.56**	20.35 (3 , 25)	0.71
2007/01	-0.11	0.56***	0.11	-5.31**	45.49 (3 , 35)	0.80
2007/02	0.24	0.39***	0.16	-8.68***	17.63 (3 , 40)	0.57
2007/03	0.31	0.46***	0.14	-7.51**	18.49 (3 , 27)	0.67
2007/04	-0.57*	0.42***	0.19**	-5.97***	28.07 (3 , 40)	0.68
2007/05	0.55*	0.49***	0.04	-10.68***	22.34 (3 , 37)	0.64
2007/06	-0.40	0.36***	0.02	-7.42***	8.16 (3 , 25)	0.49
2007/07	-0.33	0.56***	0.00	-5.57**	16.92 (3 , 32)	0.61
2007/08	0.42	0.56***	-0.06	-6.70**	10.86 (3 , 30)	0.52
2007/09	-0.28	0.37***	0.24**	-4.36*	23.79 (3 , 36)	0.66
2007/10	-0.28	0.64***	-0.03	-6.04**	11.11 (3 , 37)	0.47
2007/11	0.16	0.59***	0.15	-3.62	25.94 (3 , 32)	0.71
2007/12	0.07	0.38**	0.22**	-6.33*	15.69 (3 , 31)	0.60
2008/01	0.08	0.49***	0.12*	-6.77***	33.68 (3 , 50)	0.67
2008/02	1.14**	0.65***	0.05	-8.61**	23.69 (3 , 33)	0.68
2008/03	-0.20	0.58***	0.08	-3.42	23.95 (3 , 42)	0.63
2008/04	0.48	0.56***	0.19	-6.87**	42.43 (3 , 33)	0.79
2008/05	0.13	0.62***	0.16**	-5.39***	53.75 (3 , 53)	0.75
2008/06	0.47	0.61***	0.05	-8.26***	34.88 (3 , 56)	0.65

Each line represents the results of a separate regression of the indifference threshold on the set of explanatory variables reported in columns. **a)** Company size is measured by 11 company size classes according to the number of employees: (1/49 = 1) (50/99 = 2) (100/149 = 3) (150/199 = 4) (200/249 = 5) (250/349 = 6) (350/499 = 7) (500/649 = 8) (650/999 = 9) (1000/1999 = 10) (2000/200000 = 11).

Table 3.18. Regression of the Business Expectations Upper Indifference Threshold (USU) on Company Size and the Individual Level of Business Situation and Business Expectations Measured by the VAS

Month	Company size[a)]	Business situation	Business expect.	Constant	F	R^2
2005/10	0.08	0.05	0.72***	2.29	89.04 (3 , 83)	0.76
2005/11	-0.03	0.05	0.60***	4.74***	105.78 (3 , 92)	0.78
2005/12	-0.26	0.14***	0.58***	6.79***	70.28 (3 , 90)	0.70
2006/01	0.15	0.07*	0.65***	4.03***	148.77 (3 , 117)	0.79
2006/02	-0.04	0.08***	0.67***	4.07***	178.03 (3 , 153)	0.78
2006/03	0.21	0.07*	0.62***	3.36**	84.21 (3 , 152)	0.62
2006/04	0.29*	0.04	0.66***	3.01**	140.19 (3 , 147)	0.74
2006/05	0.38**	0.08**	0.65***	2.11*	136.7 (3 , 163)	0.72
2006/06	-0.11	0.06*	0.71***	4.62***	132.4 (3 , 136)	0.74
2006/07	-0.23	0.02	0.78***	4.38***	193.27 (3 , 147)	0.80
2006/08	-0.17	0.06*	0.73***	4.26***	147.09 (3 , 135)	0.77
2006/09	0.09	0.10**	0.74***	1.57	138.51 (3 , 110)	0.79
2006/10	0.06	0.09*	0.68***	3.27**	145.44 (3 , 113)	0.79
2006/11	0.12	0.04	0.82***	1.02	146.22 (3 , 108)	0.80
2006/12	-0.15	0.05	0.83***	3.19***	264.22 (3 , 122)	0.87
2007/01	-0.11	0.07*	0.75***	4.33***	235.23 (3 , 153)	0.82
2007/02	-0.11	0.02	0.73***	4.95***	222.75 (3 , 156)	0.81
2007/03	-0.09	0.09**	0.78***	3.08**	220.07 (3 , 157)	0.81
2007/04	-0.04	0.08*	0.70***	4.19***	219.54 (3 , 157)	0.81
2007/05	0.01	0.08**	0.69***	3.40***	218.38 (3 , 148)	0.82
2007/06	-0.05	0.02	0.83***	2.34*	215.44 (3 , 136)	0.83
2007/07	0.04	0.06*	0.80***	2.32*	269.19 (3 , 146)	0.85
2007/08	0.02	0.05	0.76***	2.73**	211.12 (3 , 136)	0.82
2007/09	-0.20	0.06	0.76***	4.44***	175.85 (3 , 136)	0.80
2007/10	-0.08	0.04	0.83***	1.74	208.58 (3 , 116)	0.84
2007/11	-0.09	0.03	0.79***	3.48***	228.25 (3 , 107)	0.86
2007/12	-0.09	0.12***	0.66***	5.65***	170.33 (3 , 97)	0.84
2008/01	0.21	0.09**	0.72***	2.82**	186.6 (3 , 132)	0.81
2008/02	0.01	0.05	0.79***	3.02**	208.22 (3 , 123)	0.84
2008/03	-0.23	0.03	0.75***	4.75***	186.01 (3 , 125)	0.82
2008/04	-0.13	0.13***	0.64***	4.48***	205.74 (3 , 145)	0.81
2008/05	0.12	0.12***	0.67***	3.03**	156.97 (3 , 116)	0.80
2008/06	-0.46**	0.05	0.74***	4.78***	116.82 (3 , 110)	0.76

Each line represents the results of a separate regression of the indifference threshold on the set of explanatory variables reported in columns. **a)** Company size is measured by 11 company size classes according to the number of employees: (1/49 = 1) (50/99 = 2) (100/149 = 3) (150/199 = 4) (200/249 = 5) (250/349 = 6) (350/499 = 7) (500/649 = 8) (650/999 = 9) (1000/1999 = 10) (2000/200000 = 11).

Table 3.19. Regression of the Business Expectations Lower Indifference Threshold (DSD) on Company Size and the Individual Level of Business Situation and Business Expectations Measured by the VAS

Month	Company size[a)]	Business situation	Business expect.	Constant	F	R^2
2005/10	0.04	0.16*	0.49***	-4.55*	38.71 (3 , 62)	0.65
2005/11	0.18	0.16***	0.45***	-7.12***	47.4 (3 , 82)	0.63
2005/12	-0.12	0.13**	0.49***	-4.33**	61.04 (3 , 74)	0.71
2006/01	0.16	0.27***	0.48***	-2.9	91.27 (3 , 60)	0.82
2006/02	0.34	-0.06	0.70***	-7.33***	28 (3 , 46)	0.65
2006/03	0.11	0.02	0.65***	-5.35*	38.69 (3 , 54)	0.68
2006/04	0.26	0.05	0.66***	-2.45	20.47 (3 , 40)	0.61
2006/05	0.22	0.39**	0.35**	-6.41***	51.47 (3 , 54)	0.74
2006/06	0.61*	0.22***	0.45***	-6.38***	48.37 (3 , 51)	0.74
2006/07	0.13	0.06	0.70***	-2.27	37.28 (3 , 43)	0.72
2006/08	-0.05	-0.02	0.81***	-2.5	65.05 (3 , 59)	0.77
2006/09	0.39	-0.01	0.80***	-4.29*	59.43 (3 , 44)	0.80
2006/10	0.09	0.13**	0.66***	-4.92***	105.27 (3 , 60)	0.84
2006/11	-0.03	0.09*	0.62***	-4.16**	73.86 (3 , 59)	0.79
2006/12	0.43	0.05	0.65***	-4.82**	93.26 (3 , 52)	0.84
2007/01	-0.45	0.21***	0.52***	-2.47	73.56 (3 , 43)	0.84
2007/02	-0.23	0.20*	0.49***	-5.26*	19.98 (3 , 34)	0.64
2007/03	0.12	0.02	0.66***	-4.64**	83.71 (3 , 37)	0.87
2007/04	-0.29	0.07	0.66***	-3.44**	98.76 (3 , 51)	0.85
2007/05	-0.08	0.13	0.50***	-6.01**	35.99 (3 , 41)	0.72
2007/06	0.20	-0.01	0.69***	-4.42**	56.62 (3 , 45)	0.79
2007/07	0.74*	0.06	0.59***	-7.42***	35.43 (3 , 39)	0.73
2007/08	-0.02	0.09	0.63***	-4.36**	65.92 (3 , 33)	0.86
2007/09	-0.16	0.17**	0.54***	-5.09***	72.99 (3 , 50)	0.81
2007/10	0.09	0.08	0.64***	-5.70*	47.72 (3 , 44)	0.76
2007/11	-0.01	0.16**	0.53***	-4.02**	61.39 (3 , 66)	0.74
2007/12	-0.22	0.11*	0.57***	-3.48*	49.68 (3 , 56)	0.73
2008/01	0.28	-0.05	0.64***	-4.76**	31.11 (3 , 57)	0.62
2008/02	0.15	0.08	0.60***	-4.76***	55.88 (3 , 62)	0.73
2008/03	-0.10	0.16**	0.42***	-3.95**	35.3 (3 , 57)	0.65
2008/04	0.26	0.1	0.57***	-6.85***	51.61 (3 , 54)	0.74
2008/05	0.05	0.11	0.60***	-5.33***	67.04 (3 , 67)	0.75
2008/06	-0.18	0.08	0.61***	-4.27***	90.56 (3 , 102)	0.73

Each line represents the results of a separate regression of the indifference threshold on the set of explanatory variables reported in columns. **a)** Company size is measured by 11 company size classes according to the number of employees: (1/49 = 1) (50/99 = 2) (100/149 = 3) (150/199 = 4) (200/249 = 5) (250/349 = 6) (350/499 = 7) (500/649 = 8) (650/999 = 9) (1000/1999 = 10) (2000/200000 = 11).

References

Abberger, K. (2006): Qualitative business surveys in manufacturing and industrial production - What can be learned from industry branch results. *Ifo Working Paper 31*.

Adcock, R. and Collier, D. (2001): Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review*, 95, pp. 529 – 546.

Ahearn, E. P. (1997): The use of visual analog scales in mood disorders: A critical review. *Journal of Psychiatric Research*, 5, pp. 569 – 579.

Aitken, R. C. B. (1969): Measurement of feelings using visual analogue scales. *Proceedings of the Royal Society of Medicine*, 62, pp. 989 – 993.

Anderson, O. (1952): The business test of the Ifo Institute for Economic Research, Munich, and its theoretical model. *Review of the International Statistical Institute* XX, pp. 1 – 17.

APA (2004): Standards for Educational and Psychological Testing. Washington, DC: American Psychological Association.

Athale, N., Sturley, A. and Skoczen, S. (2004): A Web-compatible instrument for measuring self-reported disease activity in arthritis. *Journal of Rheumatology*, 31(2), pp. 223 – 228.

Badia, X., Monserrat, S., Roset, M. and Herdman, M. (1999): Feasibility, validity and test-retest reliability of scaling methods for health states: The visual analogue scale and the time trade-off. *Quality of Life Research,* 8(4), pp. 303 – 310.

Bandholz, H. and Funke, M. (2003): In search of leading indicators of economic activity in Germany. *Journal of Forecasting,* 22, pp. 277 – 97.

Batchelor, R. A. (1981): Aggregate expectations under stable laws. *Journal of Econometrics*, 16, pp. 199 – 210.

Batchelor, R. A. (1982): Expectations, output and inflation: the European experiment. *European Economic Review*, 17, pp. 1 – 25.

Batchelor, R. A. (1986): The psychophysics of inflation. *Journal of Economic Psychology*, 7, pp. 263 – 290.

Batchelor, R. and Dua, P. (1996): Empirical measures of inflation uncertainty: a cautionary note. *Applied Economics*, 28, pp. 333 – 341.

Batchelor R. A. and Orr, A. B. (1988): Inflation expectations revisited. *Economica,* 55, pp. 317 – 331.

Becker, S. O. and Wohlrabe, K. (2008): Micro data at the Ifo Institute for Economic Research – the "Ifo Business Survey", usage and access. *Journal of Applied Social Science Studies / Schmollers Jahrbuch*, 128(1), pp. 487 – 496.

Benner, J. and Meier, C. (2004): Prognosegüte alternativer Frühindikatoren für die Konjunktur in Deutschland. *Jahrbücher für Nationalökonomie und Statistik*, 224(6), pp. 637 – 652.

Bennet, A. (1984): Output expectations of manufacturing industry. *Applied Economics*, 16(6), pp. 868 – 879.

Bijur, P., Silver, W. and Gallagher, J. (2001): Reliability of visual analog scale for measurement of acute pain. *Academic Emergency Medicine*, 8(12), pp. 1153 – 1157.

Boero, G., Smith, J. and Wallis, K. F. (2008): Uncertainty and disagreement in economic prediction: the Bank of England Survey of External Forecasters. *Economic Journal*, 118(530), pp. 1107 – 1127.

Bolton J. E. and Wilkinson, R. C. (1998): World Federation of Chiropractic Prize-Winning Paper: First Prize - Responsiveness of Pain Scales: A comparison of three pain intensity measures in chiropractic patients. *Journal of Manipulative and Physiological Therapeutics*, 21(1), pp. 1 – 7.

Bomberger, W. A. (1996): Disagreement as a measure of uncertainty. *Journal of Money*, *Credit and Banking*, 28(3), pp. 381 – 392.

Bomberger, W. A. and Frazer, W. J. (1981): Interest rates, uncertainty and the Livingstone Data. *Journal of Finance*, 36, pp. 661 – 675

Brazier, J., Green, C., McCabe, C. and Stevens K. (2003): Use of visual analog scales in economic evaluation. *Expert Review of Pharmacoeconomics and Outcomes Research*, 3(3), pp. 293 – 302.

Breitung, J. and Jagodzinski, D. (2001): Prognoseeigenschaften alternativer Indikatoren der konjunkturellen Entwicklung in Deutschland. *Konjunkturpolitik*, 47(4), pp. 292 – 314.

Brophy, S., Hunniford, T., Taylor, G., Menon, A., Roussou, T., and Callin, A. (2004): Assessment of disease severity (in terms of function) using the Internet. *Journal of Rheumatology*, 31, pp. 1819 – 1822.

Bruine de Bruin, W., Fischhoff, B., Halpern-Felsher, B. and Millstein, S. (2000): Expressing epistemic uncertainty: It's a fifty-fifty chance. *Organizational Behavior and Human Decision Processes*, 81, pp. 115 – 131.

Butler, P. V. (1997): Linear analogue self-assessment and procrustean measurement: A critical review of visual analogue scaling in pain assessment. *Journal of Clinical Psychology in Medical Settings*, 4(1), pp. 111 – 129.

Carlson, J. A. and Parkin M. (1975): Inflation expectations. *Economica*, 42, pp. 123 – 138.

Claveria, O., Pons, E. and Ramos, R. (2007): Business and consumer expectations and macroeconomic forecasts. *International Journal of Forecasting*, 23(1), pp. 47 – 69.

Claveria, O., Pons, E. and Surinach, J. (2006): Quantification of expectations: are they useful for forecasting inflation? *Economic Issues*, 11(2), pp. 19 – 38.

Clements, M. (2007): Internal consistency of survey respondents' forecasts: Evidence based on the Survey of Professional Forecasters. Mimeo, Department of Economics, University of Warwick.

Clements, M. (2008): Consensus and uncertainty: Using forecast probabilities of output declines. *International Journal of Forecasting*, 24, pp. 76 – 86

Cohen J. and Cohen, P. (1983): Applied multiple Regression: Correlation Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Coll, AM., Ameen, J. R. M. and Mead, D. (2004): Postoperative pain assessment tools in day surgery: literature review. *Journal of Advanced Nursing*, 46(2), pp. 124 – 133.

Cook, C., Heath, F. and Thompson, R. L. (2001): Score reliability in web- or Internet-based surveys: Unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement*, 61, pp. 697 – 706.

Couper, M. P. (2001): Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly*, 64(4), pp. 464 – 494.

Couper, M. P., Tourangeau, R., Conrad, F. G. and Singer, E. (2006): Evaluating the effectiveness of visual analog scales. *Social Science Computer Review*, 24(2), pp. 227 – 245.

Cronbach, L. J. (1971): Test validation. In R.L.Thorndike (ed.). Educational measurement. Second ed.. Washington, D.C.: American Council on Education.

Cunningham A. W. F., Smith R. J. and M. R. Weale (1998): Measurement errors and data estimation: the quantification of survey data, in Applied Economics and Public Policy (eds. Begg, I. and S. G. B. Henry), Cambridge Universtiy Press.

Dasgupta, S. and Lahiri, K. (1992): A comparative study of alternative methods of quantifying qualitative survey responses using NAPM data. *Journal of Business and Economic Statistics*, 19(4), pp. 391 – 400.

DeLoach, L.J., Higgins, M.S., Caplan A.B. and Stiff J.L. (1998): The visual analog scale in the immediate postoperative period: intrasubject variability and correlation with a numeric scale. *Anesthesia and Analgesia*, 86(1), pp.102 – 106.

Doepke, J. and Fritsche, U. (2006): When do forecasters disagree? An assessment of German growth and inflation forecast dispersion. *International Journal of Forecasting*, 22(1), pp. 125 – 135.

Dominitz, J. (2001): Estimation of income expectations models using expectations and realization data. *Journal of Econometrics*, 102(2), pp. 165 – 195.

Dominitz, J. and Manski, C. F. (1997): Using Expectations Data to Study Subjective Income Expectations? *Journal of the American Statistical Association*, 92(439), pp. 855 – 867.

Dominitz, J. and Manski, C. F. (2004): How should we measure consumer confidence? *Journal of Economic Perspectives*, 18(2), pp. 51 – 66.

Dreger, C., and Schumacher, C. (2005): Out-of-sample performance of leading indicators for the German business cycle. single vs combined forecasts. *Journal of Business Cycle Measurement and Analysis*, 2(1), pp. 71 – 88.

Engelberg, J., Manski, C.F. and Williams, J. (2006): Comparing the point predictions and subjective probability distributions of professional forecasters. NBER Working Paper No. W11978.

Finney, D.J. (1971): Probit Analysis. Cambridge University Press, Cambridge, UK.

Fishe, R. and Lahiri, K. (1981): On the estimation of inflationary expectations from qualitative responses. *Journal of Econometrics*, 16, pp. 89 – 102.

Foster, J. and Gregory. M. (1975): Inflation expectations: the use of qualitative survey data. *Applied Economics*, 7, pp. 319 – 329.

Freyd, M. (1923): The graphic rating scale. *The Journal of Educational Psychology*.

Fritsche, U. and Stephan, S. (2002): Leading indicators of German business cycles - An assessment of properties. *Jahrbücher für Nationalökonomie und Statistik*, 222(3), pp. 289 – 311.

Gallagher, E. J., Bijur, P. E., Latimer, C. and Silver, W. (2002): Reliability and validity of a visual analog scale for acute abdominal pain in the ED. *American Journal of Emergency Medicine*, 20(4), pp. 287 – 290.

Garson, G. D. (2007): "Reliability Analysis", from Statnotes: Topics in Multivariate Analysis. Retrieved from: http:www2.chass.ncsu.edugarsonpa765statnote.htm.

Giordani, P. and Soderlind, P. (2003): Inflation forecast uncertainty. *European Economic Review*, 47, pp. 1037 – 1059.

Gourieroux, C. and Pradel, J. (1986): .Direct Tests of the Rational Expecation Hypothesis. *European Economic Review*, 30, pp. 265 – 284.

Guiso, L. and Parigi, G. (1999): Investment And Demand Uncertainty. *The Quarterly Journal of Economics*, 114(1), pp. 185 – 227.

Guiso, L., Tiseno, A. and Winter, J. (2005): Methodological issues in the elicitation of subjective probabilities. In: A. Börsch-Supan, A. Brugiavini, H. Jürges, J. Mackenbach, J. Siegrist, G. Weber (eds.): Health, Aging, and Retirement in

Europe,. Mannheim: Mannheim Research Institute for the Economics of Aging, pp. 359 – 361.

Happich (2006): Comparing measurement methods in health state evaluation – case of the burden of Tinnitus. *Journal of Applied Social Science Studies / Schmollers Jahrbuch,* 122, pp. 351 – 367.

Hausman, J. (1978): Specification tests in econometrics. *Econometrica*, 46(6), pp. 1251 - 1271.

Henzel, S. and Wollmershaeuser, T. (2005): Quantifying inflation expectations with the Carlson-Parkin method: A survey-based determination of the just noticeable difference. *Journal of Business Cycle Measurement and Analysis*, 2(3), pp. 321 – 352.

Hüfner, F., and Schröder, M. (2002): Prognosegehalt von ifo-Geschäftserwartungen und ZEW-Konjunkturerwartungen: Ein ökonometrischer Vergleich. *Jahrbücher für Nationalökonomie und Statistik*, 222(3), pp. 316 – 336.

Ivaldi, M. (1990): Statistical methods and business surveys: An overview, in microeconometrics: Surveys and Applications (eds. J.P. Florens, M. Ivaldi, F. Laisney & J.J. Laffont), Basil Blackwell, pp. 84 – 112.

Ivaldi, M. (1992): Survey Evidence on the Rationality of Expectations. *Journal of Applied Econometrics,* 7, pp. 225 – 241.

Jensen, M. P., Chen, C. and Brugger, A. M. (2003): Interpretation of visual analog scale ratings and change scores: A reanalysis of two clinical trials of postoperative pain. *Journal of Pain,* 4(7), pp. 407 – 414.

Juster, T. (1966): Consumer buying intentions and purchase probability: An experiment in survey design. *Journal of the American Statistical Association*, 61, pp. 658 – 696.

Kahneman, D. and Tversky, A. (1979): Prospect theory: An analysis of decision under risk. *Econometrica*, XVLII, pp. 263 – 291.

Keynes, J. M. (1937): The General Theory of Employment. *The Quarterly Journal of Economics*, 51(2), pp. 209 – 223.

Klein, P. (1995): Die Konjunkturindikatoren des NBER – Measurement without Theory?, in Konjukturindikatoren (ed. Oppenländer, K. H.), München, Wien: Springer Verlag, pp. 33 – 44.

Knight, F.H. (1921): Risk, Uncertainty, and Profit. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.

Lahiri, K. and Liu, F. (2006): Modeling multi-period inflation uncertainty using a panel of density forecasts. *Journal of Applied Econometrics*, 21(8), pp. 1199 – 1219.

Lenert, L. (2000): The reliability and internal consistency of an Internet-capable computer program for measuring utilities. *Quality of Life Research*, 9, pp. 811 – 817.

Linden, S. (2003): Assessment of GDP forecasts uncertainty. *European Commission Economic Papers,* No. 184.

Löffler, G. (1999): Refining the Carlson-Parkin Method. *Economic Letters,* 64, pp. 167 – 171.

Lucas, R. (1972): Expectations and the Neutrality of Money. *Journal of Economic Theory,* 4, pp. 103 – 124.

Mankiw, G., Reis, R. and Wolfers, J. (2003): Disagreement about inflation expectations. *National Bureau of Economic Research Working Paper,* No. 9796.

Manski, C. F. (2004): Measuring expectations. *Econometrica*, 72(5), pp. 1329 – 1376.

McFadden, D. (1999): Rationality for Economists. *Journal of Risk and Uncertainty,* 19(1-3), pp. 73 – 105.

Messick, S. (1989): Validity. in R. Linn, ed., Educational measurement. New York: American Council on Education and Macmillan Publishing Company, pp. 13 – 103

Mitchell, J., Mouratidis, K. and Weale, M. (2005): Uncertainty in UK manufacturing: evidence from qualitative survey data. *Economics Letters*, 94(2), pp. 245 – 252.

Mitchell, J., Smith, R. and Weale, M. (2002a): Aggregate versus disaggregate survey-based indicators of economic activity. NIESR Discussion Paper No. 194.

Mitchell, J., Smith, R.and Weale, M. (2002b): Quantification of qualitative firm-level survey data. *Economic Journal*, 112, pp. 117 – 135.

Muth, J. F. (1961): Rational Expectations and the Theory of Price Movements. *Econometrica,* 29, pp. 315 – 335.

Nardo, M. (2003): The quantification of qualitative survey data: A critical assessment. *Journal of Economic Surveys*, 15(5), pp. 645 – 668.

Nardo, M. and Cabeza-Gutes, M. (2005): Rational expectations and the quantification of survey data: The role of measurement error. *International Journal of Applied Economics and Econometrics*, 13(3), pp. 313 – 340.

Nerlove, M. (1983): Expectations, plans, and realizations in theory and practice. *Econometrica*, 51 (5), pp. 1251 – 1280.

Nunnally, J.C. (1967): Psychometric Theory. New York: McGraw Hill, Inc.

OECD (2003): Business Tendency Surveys. A Handbook. Paris.

Parkin, D. and Devlin, N. (2006): Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Economics*, 15, pp. 653 – 664.

Pesaran, M.H. (1984): Expectations Formation and Macroeconomic Modeling, in Contemporary Macroeconomic Modelling (eds. Malgrange, P. and Muet, P.), Oxford: Blackwell.

Pesaran, M.H. and Weale, M. (2006): Survey expectations, in G. Elliott, C. W. J. Granger and A. Timmermann (eds.): Handbook of Economic Forecasting, Amsterdam: Elsevier.

Peterson, R. A. (1994): A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research,* 21(2), pp. 381 – 391.

Pfanzagl, J. (1952): Zur Methodik des Konjunkturtest-Verfahrens. *Statistische Vierteljahresschrift*, 5, pp. 161 – 173.

Ramey, G. and Ramey, V. A. (1995): Cross-country evidence on the link between volatility and growth. *American Economic Review,* 85, pp. 1138 – 1151.

Reips, U.-D. and Funke, F. (2008): Interval level measurement with visual analogue scales in Internet-based research: VAS Generator. *Behavior Research Methods* (in press).

Rich, R. and Tracy J. (2006): The relationship between expected inflation, disagreement, and uncertainty: evidence from matched point and density forecasts. Staff Reports 253, Federal Reserve Bank of New York.

Ronning, G. (1990): The Informational Content of Responses from Business Surveys, in Microeconometrics: Surveys and Applications, (eds. Florens, J.P., Ivaldi, M., Laisney F. and Laffont J.J.), Basil Blackwell, pp. 123 – 144.

Ronning, G. (1984): Wie viel Information enthält die Antwortkategorie "unverändert" in Tendenzbefragungen? *Ifo-Studien,* 30, pp. 261 – 71.

Ronning, G. (1986): Econometric Approaches to the Estimation of indifference Intervals in Business Tendency Surveys, in Business Cycle Surveys in the Assessment of Economic Activity, (eds. Oppenländer K.H. and Poser G.), Brookfield, VT: Gower.

Seitz, H. (1988): The estimation of inflation forecasts from business survey data. *Applied Economics*, 20(4), pp. 427 – 438.

Sepulveda-Umanzor, J. (2003): The relation between macro-economic uncertainty and the expected performance of the economy. Econometric Society 2004 Latin American Meeting, No 304.

Shrout, P. E. and Fleiss, J. L. (1979): Intra-class correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86, pp. 420 – 428.

Stangl, A. (2007): Ifo World Economic Survey Micro Data. *Journal of Applied Social Science Studies / Schmollers Jahrbuch,* 127(3), pp. 487 – 496.

Stangl, A. (2007): Internet Business Tendency Surveys. *Journal of Business Cycle Measurement and Analysis*, 3(3), pp. 387 – 400.

Svensson, E. (2000): Comparison of the quality of assessments using continuous and discrete ordinal rating scales. *Biometrical Journal*, 42(4), pp. 417 – 434.

Theil, H. (1955): Recent experiences with the Munich business test: An expository article. *Econometrica*, 23(2), pp. 184 – 192.

Theil, H. (1952): On the time shape of economic microvariables and the Munich business test. *Review of the International Statistical Institute XX*, pp. 105 – 120.

Tourangeu, R., Rips, L. J. and Rasinski, K. (2000): The psychology of survey response. New York: Cambridge University Press.

Tversky, A. and Kahneman, D. (1974): Judgment under uncertainty: Heuristics and biases. *Science,* 185, pp. 1124 – 31.

Tversky, A. and Kahneman, D. (1981): The Framing of Decisions and the Psychology of Choice. *Science,* 211, pp. 453 – 458.

Van Schaik, P. and Ling, J. (2003): Using on-line surveys to measure three key constructs of the quality of human-computer interaction in web sites: psychometric properties and implications. *International Journal of Human-Computer Studies*, 59(5), pp. 545 – 567.

Wagner D.R., Tatsugawa K. and Parker D. (2007): Reliability and utility of a visual analog scale for the assessment of acute mountain sickness. *High Altitude Medicine and Biology*, 8(1), pp. 27 – 31.

Westen, D. and Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology*, 84, pp. 608 – 618.

Williamson A. and Hoggart B. (2005): Pain: a review of three commonly used pain rating scales. *Journal of Clinical Nursing*, 14(7), pp. 798 – 804.

Wohlrabe, K. and Robinzonov, N. (2008): Freedom of choice in macroeconomic forecasting: An illustration with German industrial production and linear models, ifo Working Paper Nr. 57.

Wooldridge, J. M. (2002): Econometric Analysis of Cross Section and Panel Data. Cambridge, Massachusetts: The MIT Press.

Wuyts, F. L., De Bodt, M. S.and Van de Heyning, P. H. (1999): Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *Journal of Voice*, 13(4), pp. 508 – 517.

Zarnowitz, V. (1984): The accuracy of individual and group forecasts from business outlook surveys. *Journal of Forecasting*, 3, pp. 11 – 23.

Zarnowitz, V. and Lambros, L.A. (1987): Consensus and uncertainty in economic prediction. *Journal of Political Economy*, 95, pp. 591 – 621.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

_____

München, den 22. September 2008

Lebenslauf

Persönliche Angaben

Name:	Anna Stangl
Geburtsdatum:	12.09.1975
Geburtsort:	Moskau, Russland
Staatsangehörigkeit:	Deutsch

Schulischer Werdegang

09/1982 – 03/1988	Grundschule, Moskau, Russland
04/1988 – 07/1991	Gesamtschule, Erfurt
08/1991 – 06/1996	Albert-Schweizer Gymnasium, Alsfeld Abschluss: Allgemeine Hochschulreife

Akademischer Werdegang

11/1996 – 11/1997	Studium der Betriebswirtschaftslehre Friedrich-Alexander Universität, Erlangen-Nürnberg
01/1997 – 01/1998	Studium der russischen Sprache Institut für Fremdsprachen und Auslandskunde an der Universität Erlangen-Nürnberg Abschluss: Staatlich geprüfte Übersetzerin in der russischen Sprache
11/1997 – 9/2002	Studium der Sozialwissenschaften Friedrich-Alexander Universität, Erlangen-Nürnberg Abschluss: Diplom Sozialwirtin
08/2000 – 05/2001	International Student National University of Ireland, Galway, Irland
03/2005 – 01/2009	Promotionsstudim Ludwig-Maximilian Universität, München und ifo Institut für Wirtschaftsforschung